# ARGUMENT AND COMPUTATION

Series: STUDIES IN LOGIC, GRAMMAR AND RHETORIC 23(36)

Under the Auspices of the Polish Association
for Logic and Philosophy of Science

# ARGUMENT AND COMPUTATION

edited by
Marcin Koszowy

**University of Białystok**
**Białystok 2011**

**Refereed by Robert Kublikowski, Dariusz Surowik, and Robert Milewski**

Series: STUDIES IN LOGIC, GRAMMAR AND RHETORIC 23(36)
http://logika.uwb.edu.pl/studies/

Subseries issued by the Chair of Logic, Informatics and Philosophy of Science
Faculty of History and Sociology
University of Białystok



Cover design: Krzysztof Tur
Type-setting: Stanisław Żukowski

ISBN 978–83–7431–305–6
ISSN 0860–150X

WYDAWNICTWO UNIWERSYTETU W BIAŁYMSTOKU
15-097 Białystok, ul. Marii Skłodowskiej-Curie 14, tel. 857457120
http://wydawnictwo.uwb.edu.pl, e-mail: ac-dw@uwb.edu.pl

Druk i oprawa: „QUICK-DRUK” s.c., Łódź

## CONTENTS

**Marcin Koszowy**
University of Białystok

# PREFACE: KEY STRATEGIES TO ADDRESS ARGUMENT AND COMPUTATION

The problems lying at the intersection between argumentation theory and computer science constitute the subject of an intensive inquiry undertaken within the recent study of reasoning and argument. The label "Argument and Computation" characterizes the field of inquiry undertaken by the nascent research movement which has developed during the past decade.[1] The development of this movement may be illustrated by the growing activity of numerous research groups, the establishment of specialized journals, and the increasing number of monographs, conferences and workshops. Some logicians, argumentation theorists and computer scientists working in this area (see e.g. Walton & Godden, 2006; Reed & Grasso, 2007) highlight the fact that the inquiry into the overlap between argumentation theory and computer science is mutually beneficial for both disciplines:

- on the one hand, argumentation theory has brought valuable insights into the nature and structure of common sense reasoning; those insights turned out to be particularly important for building models of defeasible reasoning in Artificial Intelligence (see e.g. Rahwan & Simari, 2009);
- on the other hand, computer science, as applied to the study of argument, provided a wide range of software tools that are implemented in analyzing the structure of arguments; the key procedures which are particularly useful in accomplishing such tasks are recognizing typical argumentation schemes (see Walton, Reed & Macagno, 2008) and applying argument diagrams as tools of representing the structure of arguments (see e.g. Reed, Walton & Macagno, 2007).

[1] *Argument and Computation* is the name of the journal published by Taylor & Francis. The first issue appeared in 2010. For the motivation of the journal see (Grasso et al., 2010).

ISBN 978–83–7431–305–6 ISSN 0860-150X 

The present editorial initiative is a step towards publishing the series of volumes of the journal *Studies in Logic, Grammar and Rhetoric* devoted to the major research areas in the current study of argumentation. The first volume of this kind appeared in 2009 under the title *Informal Logic and Argumentation Theory* (vol. 16(29)). It was aimed at sketching the map of major research initiatives and approaches to argument from the 1970s to this day. This journal issue intends to give a representative sample of crucial strategies of an inquiry into the intersection between argumentation theory and computer science. Among other tasks, it discusses the implementation of formal-logical tools in representing and analyzing the structure of arguments. Such tools constitute a keystone for building computational models of argument, which are indispensable in designing computer programs employed in argument diagramming and agent communication scenarios in Artificial Intelligence. The models of argument are also discussed in the broader context of applying argumentation theory and computer science in analyzing social discourse.

In order to realize the tasks of this special issue, as sketched above, the papers of the volume discuss:

- the state of the art of inquiry into the overlap between argumentation theory and computer science;
- the applications of the systems of logic in building tools for argument analysis and evaluation;
- the implementation of argumentation systems (such as Carneades) in the study of Artificial Intelligence;
- the implementation of some ontologies for argument (such as Argument Interchange Format) as instruments providing a universal language that allows unifying various approaches to argument;
- the tools (such as model checker Perseus) for measuring the quality of persuasion dialogs;
- deductive and defeasible inference rules;
- argument schemes and diagrams;
- Internet as an instrument of argument interchange.

The authors represent major research centres and communities focusing on the study of argument. Among the contributors there are the representatives of:

- the Centre for Research in Reasoning, Argumentation and Rhetoric (CRRAR), University of Windsor, Canada;
- the Amsterdam School of Pragma-Dialectics, Department of Speech Communication, Argumentation Theory and Philosophy, University of Amsterdam and the International Learned Institute for Argumentation Studies (ILIAS), Amsterdam;

- Argumentation Research Group (ARG), School of Computing, University of Dundee, Scotland;
- the research group Argumentation, Décison, Raisonnement, Incertitude et Apprentissage (ADRIA), Institut de Recherche en Informatique de Toulouse (IRIT), Toulouse, France;
- Labóratorio de Argumentaçao (Arg Lab), Institute for the Philosophy of Language (IFL), Universidade Nova de Lisboa, Portugal;
- the PERSEUS research group (Persuasiveness: Studies on the Effective Use of Arguments), University of Cardinal Stefan Wyszyński in Warsaw and Białystok University of Technology, Poland;
- Group of Logic, Language and Information (GLLI), Opole University, Poland;
- Institute of Computer Science, Warsaw University of Technology, Poland;
- Faculty of Law, University of Białystok, Poland;
- Chair of Logic, Informatics and Philosophy of Science, University of Białystok, Poland.

The papers of the volume point to two major problems:

1. what kinds of formal tools are applied in designing computational models of argument?
2. what kinds of tools of argumentation theory are employed in representing the structure of everyday arguments?

The overview of the research field lying at the intersection between argumentation theory and computer science is presented in the paper authored by Chris Reed and Marcin Koszowy. The paper discusses the origins of the research movement, main research centers, nascent communities, monographs, articles, dedicated journals, research grants, and the possible directions of the further development of the community. The article highlights the relationship between the efforts towards building computational models of argument and the logical studies carried out in the tradition of the Lvov-Warsaw School (LWS) – the Polish philosophical movement which flourished between 1918 and 1939. Some similarities between the two traditions are exemplified by the case of MIZAR – the natural deduction system of Multi-Sorted predicate logic with Equality (MSE) which simulates the language of proofs in a simplified and standardized form, adjusted to computer processing.

The paper authored by Floris Bex and Chris Reed constitutes a systematic account of the applications of the Argument Interchange Format (AIF) – a common ontology for argument – in representing various structures of arguments. One of the goals of this research is to include within the computational model of argument not only deductive inference schemes, but also the defeasible ones. This part of the work is of crucial importance in modeling natural language arguments, in which defeasible inferences are performed. The paper discusses the applicability of argumentation scheme theory as a tool which allows taxonomizing and classifying typical patterns of reasoning. Some analyses are based upon Henry Prakken's observation that some argumentation schemes are in fact generalized inference rules (see Prakken 2010). As given examples show, the AIF is an efficient tool for representing schemes of: (a) inference (such as Defeasible Modus Ponens or Witness Testimony), (b) conflict, and (c) preference.

In the next article which is also devoted to taxonomizing arguments, Kazimierz Trzęsicki puts forward a classification of arguments upon which the method for designing argument diagrams is built. The development of Information and Communication Technologies and their implementation in Artificial Intelligence is considered as a stimulus for applying some formal tools in the study of arguments expressed in natural language. The proposed account of argument as a pair of nonempty sets of propositions embraces the intuitive notion of argument involved in natural language discourse. This approach to argument constitutes a point of departure for proposing the classification of arguments. Propositions are characterized by their relation to a system of knowledge. The types of relations between the sets and the type of propositions being the members of the sets constitute a basis for classifying arguments. Three main relations are discussed: direction of argumentation, direction of entailment, and direction of justification. Classification of arguments constitutes the groundwork for representing a variety of natural language arguments by means of argumentation diagrams. The introduced method of argument diagramming is an efficient tool in grasping various kinds of inferences, e.g. deductive, inductive, and analogical.

Another set of instruments for representing arguments are formal models of persuasive communication. The following two articles are dedicated to the applicability of formal tools in analyzing and evaluating persuasion dialogs. Leila Amgoud and Florence Dupin de Saint Cyr examine the quality of dialogs, the goal of which is persuading agents to change their minds on a given state of affairs. Three types (*families*) of criteria for evaluating persuasion dialogs are proposed: (1) measures of the quality of arguments, (2) measures concerning the components of agent's behavior (such as *coherence*, *aggressiveness* and the *novelty* of arguments), (3) measures of the quality of the dialog; the discussed criteria of evaluating a dialog's quality are *relevance* and *usefulness* of dialog moves. For each type of a persuasion dialog, the *ideal dialog* is computed. The ideal dialog is conceived as a concise sub-dialog. The quality of a given persuasion dialog is the higher the closer it is to its ideal sub-dialog.

The article authored by Katarzyna Budzyńska and Magdalena Kacprzak is another attempt at modeling persuasion dialogs formally. Persuasion dialog – a typical kind of inter-agent persuasive communication – starts with a conflict of opinion. The goal of resolving the conflict of opinion is to cause the change of agents' beliefs or commitments. The model checking technique is applied to examine the main properties of inter-agent persuasive communication. A Logic of Actions and Graded Beliefs $\mathcal{AG}_n$ is discussed as a basis upon which the model checker Perseus was designed. The authors examine the applications of Perseus in the semantic verification of $\mathcal{AG}_n$ formulas. Two kinds of procedures are performed by the system: (a) the system checks if a given $\mathcal{AG}_n$ formula is true in a given model (the standard model checking method); (b) the system searches for answer to a question concerning a given property of persuasion in a multi-agent system (the parametric verification method).

The next two contributions to the volume are devoted to the applicability of the Carneades Argumentation System in argument analysis. Carneades is an Open Source argumentation software application and library, which is employed, amongst other tasks, in argument construction with OWL ontologies and defeasible rules, calculating the acceptability of conclusions, argument mapping and visualization, goal selection, and argument interchange in XML using the Legal Knowledge Interchange Format (LKIF) (see e.g. Gordon & Ballnat, 2010). In his paper, Douglas Walton applies Carneades in the study of refuting arguments. The system is utilized to analysing cases of argument attack, challenge, critical questioning, and rebuttal. The paper clarifies the meaning of such terms as 'attack', 'rebuttal', 'refutation', 'challenge', 'defeater', 'undercutting defeater', 'rebutting defeater', 'exception', and 'objection'. A seven step procedure for seeking a refutation or objection is introduced.

The paper authored by Paweł Łoziński also contains the idea of applying Carneades in argument analysis. After giving a characteristic of Carneades, the author proposes a method of incremental analysis of arguments. Incremental analysis is confronted with argument analysis within Carneades. Whereas the method employed within Carneades relies on the search for arguments pro and con the given goal and building argumentation graph, the method of incremental argument analysis proposed by Łoziński is based on the search algorithm for choosing the exploration paths. The rationale for introducing the new method of argument analysis is given.

Edward Bryniarski, Zbigniew Bonikowski, Jacek Waldmajer, and Urszula Wybraniec-Skardowska postulate protocols concerning information networks, real interactivity systems and administering knowledge in such systems. Within the proposed account, protocols define the rules of building real dynamic epistemic logics and approximated semantics for these logics. This task is realized by employing epistemic operators related to types of communicating acts. The logical relationships related to the use of the epistemic operators are illustrated by a diagram called the square of epistemic operators. The logical relationships described within the diagram constitute the point of departure for introducing axioms for real dynamic epistemic logics. The authors extend the semantics of real dynamic epistemic logics by proposing methods of lower and upper approximation of evaluation of formulas. On the basis of those methods the 'approximation Kripke models' are defined. Some applications of the proposed tools in argument use are discussed.

The next two articles make use of Pragma-Dialects as a tradition which developed tools applicable to the inquiry into the intersection between argumentation theory and computer science. The paper authored by Jacky Visser, Floris Bex, Chris Reed, and Bart Garssen is the result of cooperation between the researchers from the Amsterdam School of pragma-dialectics and the Argumentation Research Group (ARG) (University of Dundee). It offers an original connection of two kinds of tools of argument analysis and evaluation, i.e., the Argument Interchange Format (AIF) designed by the representatives of the ARG and the pragma-dialectical model of critical discussion developed by the Amsterdam School. The pragma-dialectical model of argumentation has found so far numerous applications in the various branches of inquiry into language, reasoning and argument. The authors seek for another significant application of this model, which has not been systematically examined yet. The formalized approach to the pragma-dialectical model of a critical discussion is introduced. This account is in accord with the core research in the intersection between argumentation theory and computer science, which is of particular importance for the research in Artificial Intelligence. In order to deal with arguments computationally, at least part of models of arguments needs to be represented by means of the formal tools. The paper treats the pragma-dialectal model as a point of departure for designing a dialogue protocol which allows agents to play out a dialectical game in order to test the tenability of one agent's standpoint. Within the proposed account, the AIF allows the translation of a dialogue protocol in terms of its core ontology. The core ontology provides a directed graph data structure which allows for representing arguments. The AIF is treated as a universal language unifying various argumentation frameworks. Twofold benefits of this approach are indicated: (a) the possibility of building a normative natural language discussion model; (b) the possible implementation of the formal approach to the pragma-dialectical discussion model in an inquiry into the overlap between argumentation theory and Artificial Intelligence.

In the article which combines the tradition of pragma-dialectics with computer science, Marcin Lewiński introduces the concept of dialectical trade-offs in an argumentative discourse. Dialectical trade-offs are defined as clashes between different dialectical rules stipulated in the ideal models of argumentation, that arise in actual circumstances. The paper provides methods of dealing with the dialectical trade-offs in designing protocols for computer-mediated deliberation. The paper gives reasons for placing dialectical trade-offs on the map of the crucial fields of inquiry into the overlapping fields of argumentation theory and computer science. Lewiński makes use of the key concepts elaborated within the pragma-dialectical model of critical discussion, in particular the concept of strategic manoeuvring in an argumentative discourse. Derailments of strategic manoeuvring are discussed in terms of the choice between the good and the bad. In the context of applying the language and methods of pragma-dialectics, the nature of dialectical trade-offs is examined. Finally, loose protocols vs. formal systems for computer-aided argumentation are discussed. The proposed account of dialectical trade-offs is designed as a new tool which allows identifying and eliminating dialectical trade-offs spotted within the internet discussion forums.

The transformations of the methods of discussion in the network society are discussed by Karolina Stefanowicz, who delves into the topic of the impact of information technology on the communication process. In particular, social media are examined in terms of the new networking tools. Possible applications of the 20th century philosophical conceptions of public sphere in developing methods of analysing new tools for social communication are considered. The author characterizes the consequences of using main tools of the new social dialogue and the consequences of its use. The opportunities and threats of applying new tools of communication are examined.

From what has been presented above, the efforts of joining various research perspectives and approaches to argument and reasoning are noticeable within the recent strands of inquiry into the overlap between argumentation theory and computer science (esp. Artificial Intelligence). I owe special thanks to Chris Reed, Robert Kublikowski, Rafał Lizut, Kazimierz Trzęsicki, Dariusz Surowik, and Ewa Wasilewska-Kamińska for their valuable comments on this volume.

## References

Besnard P., & Hunter, A. (2008). *Elements of Argumentation*. Cambridge, Mass. & London: The MIT Press.

Gordon, T. F., & Ballnat, S. (2010). The Carneades Argumentation System. COMMA 2010 Conference Website, http://www.ing.unibs.it/˜comma 2010/demos/.

Grasso, F., Rahwan, I., Reed, C., & Simari, G.R. (2010). Editorial. *Argument and Computation, 1(1)*, 1–5.

Prakken, H. (2010). On the nature of argument schemes. In C. Reed & C. Tindale (Eds.), *Dialectics, Dialogue and Argumentation. An Examination of Douglas Walton's Theories of Reasoning and Argument* (pp. 167–185). London: College Publications.

Rahwan, I., & Simari, G. R., (2009). Preface. In. I. Rahwan & G. R. Simari (Eds.), *Argumentation in Artificial Intelligence* (pp. IX–X). Dordrecht etc.: Springer.

Reed, C., & Grasso, F. (2007). Recent Advances in Computational Models of Argument. *International Journal of Intelligent Systems, 22(1)*, 1–15.

Reed, C., Walton, D., & Macagno, F. (2007). Argument diagramming in logic, law and artifical intelligence. *The Knowledge Engineering Review, 22*, 87–109. doi:10.1017/S0269888907001051.

Walton, D., & Godden, D.M. (2006). The impact of Argumentation on Artificial Intelligence. In P. Houtlosser & A. van Rees (Eds.), *Considering Pragma-Dialectics* (pp. 287–299). Mahwah, NJ: Lawrence Erlbaum.

Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation Schemes*. Cambridge etc: Cambridge University Press.

Marcin Koszowy
Chair of Logic, Informatics and Philosophy of Science
University of Białystok, Poland
koszowy@uwb.edu.pl

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 23 (36) 2011

**Chris Reed**
University of Dundee

**Marcin Koszowy**
University of Białystok

# THE DEVELOPMENT OF ARGUMENT AND COMPUTATION AND ITS ROOTS IN THE LVOV-WARSAW SCHOOL

**Abstract**: The paper discusses the relation between computational models of argument and the study of reasoning carried out within the tradition of the Lvov-Warsaw School (LWS). Section 1 presents the origins and the recent strands of inquiry into the overlap between argumentation theory and computer science. Section 2 refers to the legacy of the study of reasoning in the Lvov-Warsaw School. Some research areas of the School which correspond to the contemporary study of argument and computation are indicated. Reasons for applying methods of automated reasoning (esp. the MIZAR system) in argument analysis are given.

**Keywords**: argument, computation, Lvov-Warsaw School (LWS), computer-assisted reasoning, MIZAR

## 1. Argument & Computation

### 1.1. The domain and community

Over the past ten years or so, a new interdisciplinary field has emerged in the ground between, on the one hand, computer science – and artificial intelligence in particular – and, on the other, the area of philosophy concentrating on the language and structure of argument. There are now hundreds of researchers worldwide who would consider themselves a part of this nascent community. Various terms have been proposed for the area, including "Computational Dialectics", "Argumentation Technology" and "Argument-based Computing", but the term that has stuck is simply Argument and Computation. It encompasses several specific strands of research:

- the use of theories of argument, and of dialectic in particular, in the design and implementation of protocols for multi-agent communication;
- the application of theories of argument and rhetoric in natural language processing and affective computing;

- the use of argument-based structures for autonomous reasoning in artificial intelligence, and in particular, for defeasible reasoning;
- computer supported collaborative argumentation – the implementation of software tools for enabling online argument in domains such as education and e-government.

These strands come together to form the core of a research field that covers parts of AI, philosophy, linguistics and cognitive science, but, increasingly is building an identity of its own. The diversity of research conducted in Argument and Computation reflects the different disciplinary points of origin, including:

- formal models of argumentation systems, originating in the nonmonotonic reasoning community;
- argumentation in legal reasoning, originating in the AI and Law community;
- the language of argument, originating in the discourse analysis and corpus linguistics communities;
- argument in multi-agent systems, originating in the distributed AI community;
- computer supported collaborative argumentation and argument visualisation, originating in the computer supported collaborative work community;
- argumentation-based pedadogy, originating in the AI and Education community;
- probabilistic argumentation, originating in the Bayesian reasoning community;

and covers many different specific themes, including:

- Argumentation and cognitive architectures;
- Argumentation and computational game theory;
- Argumentation and defeasible reasoning;
- Argumentation and nonmonotonic logics;
- Argumentation and Decision Theory;
- Argumentation and Logic Programming;
- Argumentation and game semantics;
- Software for teaching argumentation skills;
- Argumentation-based interaction protocols;
- Argumentation-based semantics of programs;
- Argumentation in natural language processing;
- Argumentation in human computer interaction;
- Argumentation in multi-agent systems;
- Computational models of natural argument;

- Dialogue games and conversation policies;
- Dispute resolution and mediation systems;
- Electronic democracy and public deliberation;
- Legal and medical applications;
- Models of bargaining and economic interaction;
- Reasoning about action through argumentation;
- Computational tools for argumentation support.

The diversity of contributing backgrounds is also reflected in the geographical distribution of the work. Though catalysed largely in Western Europe, there is a broad distribution of research across the world, of which the largest groups are based:

- in Argentina at Universidad Nacional del Sur, Argentina;
- in France at the Institut de Recherche en Informatique de Toulouse;
- in Germany at Fraunhofer FOKUS Berlin;
- in Italy at Universita Degli Studi di Brescia;
- in Luxembourg at the University of Luxembourg;
- in the Netherlands at Universiteits van Amsterdam and Utrecht and Rijksuniversiteit Groningen;
- in Thailand at the Asian Institute for Technology;
- in the UAE at the Masdar Institute;
- in the UK at the Universities of Aberdeen, Dundee, Liverpool, Southampton and Imperial College London;
- in the US at the City University of New York;

but in addition to these centres – which often serve to catalyse or connect research communities nationwide – there are also vibrant argument and computation communities in Austria, Australia, Belgium, Canada, Poland, Spain, Sweden, Switzerland amongst others.

2000 represents one good point at which to mark the rise of the interdisciplinary area between computing (specifically, artificial intelligence) and argumentation. Before that, there were occasional conferences such as *Formal and Applied Practical Reasoning* (Gabbay & Ohlbach, 1996) and workshops, such as those on Computational Dialectics in the mid 90's organised by Loui, Gordon et. al. But otherwise little else.

In 2000, the Symposium on Argument and Computation brought together philosophers, AI researchers, linguistics, psychologists, lawyers and rhetoricians in a structured way to collaborate on a book project which turned out very successfully as the *Argumentation Machines* book published in Kluwer's Argumentation Library. Independently, the CSCW community was developing links with practical reasoning philosophers and educators in developing visualisation and group-working systems (see, e.g. The CSCA workshops organised by Buckingham-Shum, which resulted in (Kirschner et al., 2003)). Philosophers of argument were also starting to interact with AI independently (e.g. Walton with multi-agent systems, Hitchcock with defeasible reasoning, and Jackson with AI and education, amongst many others).

Over the following few years, there has been steady growth. The CMNA workshops, for example, organised by Grasso, Reed and latterly, Green have helped to nurture that growth since 2001 (2001, ICCS, San Francisco was a small meeting; then 2002, ECAI, Lyon was a full workshop; 2003, with IJCAI in Acapulco; 2004, with ECAI in Valencia; 2005, with IJCAI in Edinburgh; 2006 with ECAI in Riva del Garda was the first time the workshop was a 2-day event; 2007 with IJCAI in Hyderabad; 2008 with ECAI in Patras; 2009 with IJCAI in Pasadena; 2010 with ECAI in Lisbon and in 2011, with AAAI in San Francisco). Grasso and Reed produced a special issue of the International Journal on Intelligent Systems with resubmitted and re-reviewed material from the first three CMNA workshops, for which the introductory editorial provides a thorough overview of the field at that time (Reed & Grasso, 2007).

2004 also witnessed the introduction of another relevant workshop series focusing on argumentation in multi-agent systems, ArgMAS run with the AAMAS conference in New York. This workshop is co-organised each year by a subset of the steering committee comprising Kakas (Cyprus), Maude (Paris, France) McBurney (Liverpool, UK), Moraitis (Paris, France), Parsons (New York, US), Rahwan (Masdar, UAE) and Reed (Dundee, UK). It has a healthily selective acceptance rate and publishes proceedings with Springer. It is held with AAMAS every year, after New York in 2004 it was held at Utrecht in 2005, Hakodate in 2006, Hawaii in 2007, Estoril in 2008, Budapest in 2009, Toronto in 2010 and Taipei in 2011.

2006 saw the inauguration of the new international conference series on Computational Models of Argument, COMMA. The second COMMA conference was held in Toulouse in 2008, and the third in Brescia in 2010. In 2012, it will be in Vienna. The third COMMA conference saw the formal launch of the new journal dedicated to the area, the *Journal of Argument and Computation*, and this journal has been recognised by its publisher, Taylor & Francis (who use the Routledge imprint in philosophy) for its high rate of both selectivity and citations in its first few years.

The first decade of the century also saw an increasing number of journal special issues dedicated to various computational aspects of argument, covering some of the most high profile journals in the field including:

- *Computational Intelligence* (Blackwell, 2001);
- *Journal of Logic and Computation* (OUP, 2003);

- *Autonomous Agents and Multi-Agent Systems* (Springer, 2005);
- *Artificial Intelligence and Law* (2005);
- *Argumentation* (Springer, 2006);
- *International Journal of Intelligent Systems* (World Scientific, 2006);
- *Artificial Intelligence* (Elsevier, 2007);
- *IEEE Intelligent Systems* (IEEE, 2007).

Following on from the success of its special issue, the *Journal of Logic and Computation* also in 2009 introduced a special track of 'corner' on argument and computation.

Finally, there has also been a concommitant increase in funders' recognition of the importance of the area with a variety of projects across Europe and worldwide, representing, between them, over €20m of support for research into argument and computation, including:

- ASPIC (EU funded, 2004–7);
- ArgueGRID (EU funded, 2006–8);
- AMI and AMIDA (EU funded, 2004–9);
- I-Exchange (EPSRC funded, 2004–6);
- Dialectical Argumentation Machines (EPSRC funded, 2009–12);
- Argumentation Factory (EPSRC funded, 2006–9);
- ITA (DARPA funded, 2006–16).

Of course, many more national and international projects have touched upon themes in the argument and computation space as well.

### 1.2. The research of the field

It is convenient to summarise the major landmarks in the field to give an introduction to, and orientation within, the domain of argument and computation. Fuller introductions can be found in (Reed & Grasso, 2007) and (Reed & Norman, 2003b) amongst others: the aim here is simply to sketch the main advances.

An early paper outlining the role that argumentation plays in unifying particular types of logic – and in particular, nonmonotonic logics – was (Lin & Shoham, 1987), which shows how many of the major approaches (both then and now) to understanding and modelling reasoning in AI can be seen as instances of argumentation. Circumscription, default logic, nonmonotonic logic and defeasible logic were all demonstrated to be special cases of a more general argumentation-based logic, showing not only that there are strong connections between these system (which was to have been expected, but had not previously been shown formally) but also that argumentation is a powerful notion for understanding and interpreting formal computational systems.

In 1995, two major landmark papers appeared which are now considered to be foundational works. (Krause et al., 1995) describes the 'logic of argumentation', LA, which laid the foundation for a rich seam of theoretical and applied work by the British cognitive and computer scientist, John Fox and colleagues – one which continues today. (Dung, 1995) described a formal notion of 'acceptability' which allows for the development of various types of semantics of argumentation. The approach has subsequently been described elqouently (by Prakken) as a 'calculus of opposition', and has driven a small industry of research into the development of various variations, extensions and applications which again is still growing today. The same year a major landmark in the domain-specific segment of argument and computation dedicated to legal argument also appeared: (Gordon, 1995).

2000 to 2002 saw publication of several important review articles, to which the interested reader is referred for a more comprehensive treatment of particular facets of argument and computation. (Carbogim et al., 2000) review techniques for representing and reasoning with knowledge using argumentation structures; (Chesñevar & Maguitman, 2000) review logical approaches to argumentation, and (Prakken & Vreeswijk, 2002) to defeasible argumentation in particular. (McBurney & Parsons, 2002) review the area of dialogue games in multi-agent systems.

The two significant monographs in 2003, (Reed & Norman, 2003a) and (Kirschner et al., 2003) already mentioned, coincide with the rapid growth in the number of people working in the area and the related increase in recognition and citation of the work. More recently, monographs primarily within the field of philosophy have also started to appear as a result of sustained interdisciplinary collaborations (such as Walton et al., 2008).

Finally, work on the Argument Interchange Format, started in 2006 (Chesñevar et al., 2006), has begun to bring together many of the disparate techniques and approaches into a framework that supports interchange, evaluation, and resource re-use across tools and theories and represents an exciting new potential hub around which future research might be conducted.

In the context of this collection, it is also worth highlighting the recent establishment of a nascent community of scholars working on argument and computation in Poland, an effort spearheaded by Budzyńska (UKSW) and Kacprzak (Białystok) in collaboration with colleagues at PAN, and the Universities of Poznań and Warsaw, amongst others (see, for example, publications such as (Budzyńska et al., 2009; Budzyńska & Dębowska, 2010)), and the ArgDiaP series of workshops (http://www.argdiap.pl). As this national community develops its own coherence, it has started to collaborate with those internationally, resulting in publications such as (Kacprzak et al., 2007; Bex & Budzyńska, 2010; Dębowska et al., 2009).

With the research area of argument and computation now established in both the computational and philosophical communities (appearing as special tracks, themes, or sections of major conferences such as IJCAI and APA and major journals such as the Journal of Logic and Computation and Synthese), and developing an identity of its own with the COMMA conference and the Journal of Argument and Computation, the field looks set to grow in breadth and maturity, a growth to which this special issue is aimed at supporting and encouraging. The articles of the volume discuss key topics presented in this section, as well as some new lines of inqiury. Among the addressed issues there are: applications of the Carneades Argumentation System (Walton, 2011; Łoziński, 2011), formal tools for evaluating persuasion dialogues (Amgoud & Dupin de Saint Cyr, 2011; Budzyńska & Kacprzak, 2011), applications of the AIF in representing schemes of inference, conflict, and preference (Bex & Reed, 2011), argument diagrams (Trzęsicki, 2011), the implementation of epistemic logics in argument analysis and evaluation (Bryniarski, Bonikowski, Waldmajer, and Wybraniec-Skardowska, 2011), the connections between the study of argument and computation and the Pragma-Dialectical Discussion Model (Visser, Bex, Reed, & Garssen, 2011; Lewiński, 2011), and the impact of information technologies on the social discourse (Stefanowicz, 2011).

## 2. Reasoning and computation – the legacy of the Lvov-Warsaw School

### 2.1. Main research areas

The Lvov-Warsaw School (LWS) was established by Kazimierz Twardowski at the end of the 19th century in Lvov. Along with the development of logic there were systematically carried out studies in ontology, epistemology, ethics, aesthetics, methodology of science, philosophy of science, semiotics, and philosophy of language (see Woleński, 1989, Ch. 1–2; Jadacki, 2006). Among other achievements in various branches of philosophy, the school is famous for its achievements in mathematical logic (see e.g. Woleński 1989, Ch. 1, part 2). In 'the golden age of Polish logic', which lasted for two decades (1918–1939), formal logic became a kind of an 'international visiting card' of the LWS (Jadacki 2009, p. 91; see also Falkenberg 1996).[1]

[1] The key role in popularizing logical ideas of the LWS was played by Heinrich Scholz

The keystone for the developments in formal logic was laid, among many others, by Jan Łukasiewicz, Stanisław Leśniewski, Alfred Tarski, Kazimierz Ajdukiewicz, Tadeusz Czeżowski, Bolesław Sobociński, Andrzej Mostowski, Adolf Lindenbaum, Stanisław Jaśkowski, Mordechaj Wajsberg, Mojżesz Presburger, Jerzy Słupecki, and Bolesław Sobociński (see e.g. Woleński, 1995, pp. 369–378; Jadacki, 2009, pp. 11–20; Wybraniec-Skardowska, 2009, pp. 6–8).

This section is based on works of few representatives of the LWS, and on works of successors of the LWS. It aims at sketching an answer to the question: which logical ideas of the LWS may be employed in the area of building computational models of argument? Among many issues discussed within the logical studies carried out in the LWS, there are two topics which may be of interest in the context of investigating the issues on the boundary between argumentation theory and computer science:

1. the concepts of logic and reasoning – for these concepts illustrate the tendency to combine formal analysis of arguments with the pragmatic characteristics of the context of argument use;
2. the impact of some logical ideas of the LWS on computer science – for it indicates possibility of applying further the language and methods of logic to building computational models of reasoning; among these ideas there are (see Trzęsicki, 2007, pp. 19–29):
   - Polish notation (parenthesis-free notation) invented by Jan Łukasiewicz;
   - multi-valued logics also created by Łukasiewicz;
   - the system of natural deduction invented by Stanisław Jaśkowski (independently of Gerhard Gentzen);
   - discursive logic developed by Stanisław Jaśkowski;
   - impact of some ideas of Jerzy Łoś on the invention of temporal logic by Arthur Norman Prior;
   - categorial grammar developed by Kazimierz Ajdukiewicz;
   - the theory of recursive functions elaborated by Andrzej Grzegorczyk.

Since one of the goals of designing computational models of argument is developing computer-aided procedures of argument analysis, in what follows, a possible application of a system of automated reasoning in representing arguments will be given. A key idea applied in designing systems of

(1930), who is claimed to be the first modern historian of logic (Woleński, 1995, p. 363) For the discussion on Scholz's role in propagating the LWS see Jadacki, 2009 (Ch. 8: *Heinrich Scholz and the Lvov-Warsaw School*, pp. 155–171).

computer-aided reasoning is Stanisław Jaśkowski's system of natural deduction. For it constituted a theoretical inspiration for designing Mizar – the system of a computer-aided representation and verification of mathematical knowledge.[2] Therefore some applications of Mizar in argument representation will be suggested.

### 2.2. Concepts of logic and reasoning

The very concepts of logic and reasoning present in the works of the LWS representatives illustrate the tendency to combine formal analyses of reasoning with some pragmatic account of the context of reasoning. The concept of logic present in the works of some thinkers of the LWS (see e.g. Ajdukiewicz, 1974, pp. 2–4) embraces not only formal logic, but also semiotics and methodology of science. Within this broader account of logic the tendency to treat formal logic as an indispensable, but not exclusive tool of the study of reasoning has been developed. Hence, the study of reasoning in the LWS is surely not tailored for applying the formal-logical tools in analyzing and evaluating reasoning.[3]

A possible point of departure of the logical studies of argument within the tradition of the LWS[4] is conceiving an argument as a pair of nonempty sets of propositions. For example, arguments are structures $\langle \Sigma, \Gamma \rangle$, where $\Sigma$ is the set of premisses and $\Gamma$ is the set of conclusions. Among the relations between $\Sigma$ and $\Gamma$ there are: direction of argumentation, direction of entailment, and direction of justification (see Trzęsicki 2011, this issue). An example of a tendency to include pragmatic concepts (such as justification within a given context) into symbolic representations of arguments is the 'pragmatic concept of inference' which was introduced by another representative of the LWS, Seweryna Łuszczewska-Romahnowa (1962). According to this approach, the proposition $p_k$ follows pragmatically (given the theoretical context) from the sequence of propositions $p_1, \ldots, p_n$ if and only if the implication $p_1, \ldots, p_n \to p_k$ has been justified within this context. A similar approach is present in Kazimierz Ajdukiewicz's analyses of subjectively uncertain inference (1974, Ch. 4, pp. 120–181). The pragmatic account of arguments may also manifest itself through introducing prag-

---

[2] http://mizar.org.

[3] Due to the fact that reasoning was carefully investigated in the LWS (see e.g. Jadacki, 2009, pp. 98-99), classifications of reasonings were designed by the major representatives of the LWS, for example by Łukasiewicz, Czeżowski, and Ajdukiewicz (see Woleński, 1988; Kwiatkowski, 1993).

[4] It is not claimed, however, that this point of departure is specific exclusively for the logical studies carried out in the LWS.

matic predicates (such as ‘assume that’, ‘allow that’, and ‘assert that’) and the pragmatic concept of subjective (psychological) probability (Budzyńska, 2004, pp. 128–129).[5]

Another illustration of accepting a broader account of arguments within the legacy of the LWS is a general argumentation framework presented by Jan Woleński (2008, p. 105). Argumentation is examined as a sequence of moves $\alpha_1, \alpha_2, \ldots, \alpha_n \rightsquigarrow \beta$, where $\beta$ is a thesis (claim, view, standpoint), $\alpha_1, \alpha_2, \ldots, \alpha_n$ is a finite sequence of argumentative moves made in order to convince an audience to accept $\beta$, and $\rightsquigarrow$ denotes the relation of acceptance of the thesis. This general argumentation framework may be treated as a point of departure for characterizing argumentative moves from the point of view of (a) formal logic and (b) pragmatics. The main question raised from the point of view of logic is: does $\beta$ follow logically from $\alpha_1, \alpha_2, \ldots, \alpha_n$? The pragmatic approach to a given sequence of moves is based on treating them as persuasive moves of the proponent.

The next example of the pragmatic account of reasoning is Witold Marciszewski’s definition of argument as reasoning whose aim is to influence an audience:

> A reasoning is said to be an argument if its author, when making use of logical laws and factual knowledge, also takes advantage of what he knows or presumes about his audience’s possible reactions (Marciszewski, 1991, p. 45).

The remark that the knowledge about the audience’s reactions plays a key role in any successful persuasion is a point of departure for seeking theoretical foundations for the art of argument not only in formal logic, but also in accounts of human cognition and the mind-body relations, as present in philosophy and in cognitive science.[6] In what follows the basic features of this approach will be discussed.

[5] For the analysis of Ajdukiewicz’s account of the subjectively uncertain inference see (Koszowy 2010).

[6] An example of employing this broader approach is the research project *Undecidability and Algorithmic Intractability in the Social Sciences*, which was realized from 2003 to 2006 at the University of Białystok. The research was supported by the Polish Committee for R&D Ministry of Science (Grant No. 2 H01A 030 25). The project was coordinated by Witold Marciszewski. Amongst other goals, the research focused on identifying some problems that are (algorithmically) undecidable or intractable (Marciszewski, 2003, pp. 79–80; 2006a, p. 9; 2006b, pp. 145–157).

### 2.3. Logical ideas of the LWS and the computational models of argument

An example of developing an account of argument from the point of view of computing is Witold Marciszewski's approach to an argument (1991). This account is rooted in a conception of reasoning as computing, which is the most briefly expressed with Gottfried Wilhelm Leibniz's call: *Calculemus*![7] Within Marciszewski's approach, the concept of information processing constitutes a theoretical foundation of the art of argument. Information is treated as a theoretical entity recorded in a material vehicle. Two kinds of records of information are distinguished: *external* (information is not part of a communicating system) and *internal* (information is part of a communicating system). Next, two ways of information processing are distinguished: *direct processing* (performed without recording), and *indirect processing* (performed with producing records).

Those two distinctions allow answering the question: what is the place of arguments on the map of information-processing phenomena? Arguments are located in the area of indirect processing of consciousness with external records, and then in processing internal records by the corresponding acts of consciousness (Marciszewski, 1991, p. 46).

The next theoretical tool for dealing with the structure of arguments is the framework of transforming a sequence through appending new elements. Within this framework one may distinguish a sequence which belongs to a definite (1) domain. Items in that sequence are created by applying a definite (2) operation (a many-one or one-one transformation). The sequence tends to (3) a bound either in virtue of that operation itself or by our decision as to the point to stop. When generating a next element of the sequence by employing a definite operation, a trait of preceding elements is preserved – this trait is called (4) an invariant. Within this framework, arguments ruled by formal logic are characterized as follows:

1. a domain consists of propositions;
2. operations are defined by inference rules;
3. a bound is a conclusion one seeks for;
4. a preserved trait (invariant) is a logical value called truth.

[7] Leibniz's legacy is stressed by Witold Marciszewski, who is an administrator of the WWW domain "Calculemus" (www.calculemus.org). The goal of this domain is, among other tasks, to expose the impact of Leibniz's logical and philosophical ideas on the origins and development of computer science (see Marciszewski, 1997; Trzęsicki, 2007). For some results of a research project *Logical Systems and Algorithms for Automatic Testing of Reasoning* (1986-1990) concerning mechanization of reasoning see (Marciszewski & Murawski, 1995).

Propositions from the proof (i.e. premisses and conclusions) may be treated as pieces of information:

> It is difficult to articulate an adequate definition of information processing, however the practice of proving theorems provides us with a partial, at least, operational definition. It is so, because anybody who proves a proposition on the basis of other ones, thereby displays an intuitive understanding of what a proposition is, and it does represent, indeed, a typical piece of what we call information (Marciszewski, 1991, p. 47).

Hence, transforming premisses into a conclusion is treated as a paradigmatic example of information processing:

> [...] the rules of inference deal solely with graphical transformations of formulas, i.e. with changing their shapes, and at the same time abstract entities, viz. propositions, are attached to those external records. Thus the processing of the record gets matched with the processing of information corresponding to that record (likewise, operations on numbers as abstract entities correspond to the processing of digits) (ibid.).

Taking into account the explanatory power of this example of information processing, Marciszewski treats it as a heuristic model of human reasoning:

> Human thoughts (in a psychological sense), as phenomena occurring in time, together with their records in the internal language of a (biological machine) are to be construed as spatio-temporal instantiations of abstract entities being propositions (ibid).

Hence, this framework may serve as a useful heuristic model in analyzing logical fallacies by comparing deductively invalid inference schemes with this model. Since the universal laws of information processing are common to all information-processing systems (both to human beings and to computers), this model is claimed to be applicable in analyzing various information processing phenomena, despite the fundamental differences between human beings and cipher machines (p. 48). However, the discussed model is not claimed to be a unique legitimate tool for analyzing arguments, for it does not deal with defeasible inference schemes.

The main features of the proposed approach to arguments may constitute a point of departure for research projects which mainly aim at:

- placing arguments in the framework of information processing;
- analyzing arguments in terms of external records, especially of formalized proofs as a paradigm of information processing.

These goals are realized by systems for automated reasoning, automated deduction, and automated proof checking. Mizar[8] is an example of such a system. The Mizar project started in 1973, on the initiative of Andrzej Trybulec. Mizar is (1) a formal language for writing formalized mathematical definitions and proofs, (2) a computer program used for verifying mathematical proofs (see Trybulec 1993, Matuszewski & Rudnicki 2005; Grabowski et al., 2010). Since 1989 the focus of the project has been also to develop a database for mathematics (*Mizar Mathematical Library* – MML). Marciszewski (1994) describes Mizar as:

> (i) a natural deduction system of (ii) Multi-Sorted predicate logic with Equality, for short MSE, (iii) that simulates the language of proofs, esp. that used by mathematicians, in a simplified and standardized form, adjusted to computer processing, and (iv) that is combined with a proof checker, i.e. a program checking proof validity (Marciszewski, 1994).

In order to make the connections between the methods of analyzing reasoning in the legacy of the LWS and the methods of building computational models of argument more explicit, we shall discuss two main theses concerning possible applications of Mizar in proposing a kind of a computational model of argument. The theses hold that:

- the Mizar language is a useful tool of representing the structure of arguments;
- the Mizar methods of automated proof-checking are applicable in identifying formal logical fallacies.

In order to present applicability of the methods, first some basic features of the Mizar language shall be briefly discussed. Since Mizar is based on the first order predicate logic (Grabowski et al., 2010, p. 155; Wiedijk, 2011, p. 1, 50), statements are composed of atomic (predicative) formulas combined with connectives and quantifiers of classical logic. The main logical connectives and quantifiers are expressed as follows (ibid.):

| | |
|---|---|
| $\neg\alpha$ | not $\alpha$ |
| $\alpha \wedge \beta$ | $\alpha$ and $\beta$ |
| $\alpha \vee \beta$ | $\alpha$ or $\beta$ |
| $\alpha \rightarrow \beta$ | $\alpha$ implies $\beta$ |

[8] When referring to the origins of the name of system, Marciszewski (1994) states: *Don't try to guess what the name "Mizar" means. It was the author's fancy to take a star's name (...) [to stand for it].*

| | |
|---|---|
| $\alpha \Leftrightarrow \beta$ | $\alpha$ iff $\beta$ |
| $\exists_x \alpha$ | ex x st $\alpha$ |
| $\forall_x \alpha$ | for x holds $\alpha$ |
| $\forall_{x:\alpha} \beta$ | for x st $\alpha$ holds $\beta$ |

Since a type of each quantified variable has to be given, the form of quantifiers may be as follows (ibid.):

```
for x being set holds...
```

or

```
ex y being real number st...
```

In order to present a possibility of representing arguments in the MIZAR language we shall consider an example of a fallacy of affirming the consequent (AC). Let us take the following reasoning: *if one is able to make a cipher machine intelligent, then one may understand intelligence. One understands intelligence, therefore one is able to make a cipher machine intelligent.* This reasoning falls under the invalid inference scheme:

$$\frac{\begin{array}{l} p \rightarrow q \\ q \end{array}}{p}$$

The representation of the fallacy in the MIZAR style is as follows:

```
environ
begin
:: p[]
:: q[]
scheme Invalid Rule {p[],q[]}:
p[]
provided
A1:p[] implies q[]
and
A2:q[]
proof
  thus p[] by A1,A2;
::>       *4
end;
```

The system identifies the logical invalidity of this reasoning by showing the error '*4'.

This inference scheme may be contrasted with a valid inference scheme such as Modus Ponens (if p then q, and p, therefore q), which is expressed in the MIZAR style as follows:

```
environ
begin
:: p[]
:: q[]
scheme ModusPonens {p[],q[]}:
q[]
provided
A1:p[] implies q[]
and
A2:p[]
proof
  thus q[] by A1,A2;
end;
```

This time, directly after drawing the conclusion ($q$) from the premisses ($A1$, $A2$), no error occurs, because of the fact we have the logically valid inference scheme.

In order to show how predicates are expressed in MIZAR, let us consider the second example, which alludes to an imitation game upon which the Turing Test was designed. Let us imagine someone trying to guess whether his or her interlocutor is a human being or a machine. Let us now consider the following line of reasoning: *every interlocutor is either a human being or a computer, therefore either every interlocutor is a human being or every interlocutor is a computer.* This reasoning has a following deductively invalid inference scheme:

$$\frac{\forall x[P(x) \vee Q(x)]}{\forall x P(x) \vee \forall x Q(x)}$$

The reasoning may be expressed in the MIZAR style as follows. Instead of the letters 'P' and 'Q' we can also use names such as 'Human Being' and 'Computer':

```
environ
begin
reserve x for set;
scheme Ex1{HumanBeing[set],Computer[set]}:
```

```
(for x holds HumanBeing[x]) or (for x holds Computer[x])
provided
A1: for x holds HumanBeing[x] or Computer[x]
proof
  thus (for x holds HumanBeing[x])
       or (for x holds Computer[x]) by A1;
::>                                 *4
end;
```

Again, the fallacy is identified with '*4'. In order to show how valid inference may be expressed by MIZAR, let us consider the inference scheme:

$$\frac{\forall x P(x) \vee \forall x Q(x)}{\forall x [P(x) \vee Q(x)]}$$

The reasoning which is in accordance with its scheme may be expressed in MIZAR as follows:

```
environ
begin
reserve x for set;
scheme Ex2{HumanBeing[set],Computer[set]}:
for x holds HumanBeing[x] or Computer[x]
provided
A1: (for x holds HumanBeing[x]) or (for x holds Computer[x])
proof
  thus for x holds HumanBeing[x] or Computer[x] by A1;
end;
```

After the conclusion (`for x holds HumanBeing[x] or Computer[x]`) is drawn, no error occurs.

The above examples[9] illustrate the possibility of applying systems of computer-aided mathematical reasoning both in argument representation and in identification of formal fallacies. Some future inquiry into applications of MIZAR in analyzing fallacies may consist in detecting some formal logical fallacies on the basis of analyzing the structure of reasoning. This task is in accord with deductivism – the view which holds that fallacies may be identified as deductively invalid inferences (see e.g. Jacquette, 2007;

[9] We are grateful to Mariusz Giero and Karol Pąk for discussion of examples. For nontrivial examples of natural deduction proofs see (Pąk, 2010, pp. 100–105).

Jacquette, 2009). Those possible applications are also in line with some initial attempts to propose computational methods of detecting formal logical fallacies in an argumentative discourse, such as those made by Gibson, Rowe, and Reed (2007, pp. 27–29), in which an example of the fallacy of affirming the consequent is represented in the XML and in the AML (*Argument Markup Language*, based on the XML). However, this general idea of computer-aided detection of formal logical fallacies needs to be further systematically developed. Examples given in the paper show how the essential features of the Mizar language may be instructive in an inquiry into this field.

Hence, possible applications of systems of automated reasoning may be justified by indicating those twofold profits:

1. representation of argument schemes by means of a computer-aided knowledge representation enriches the palette of devices of mathematical knowledge representation;
2. expressing the structure of arguments in Mizar may be instrumental in exposing the key similarities between the project of automated reasoning and the study of computational models of natural argument.

Moreover, some key features of the Mizar language, such as clarity in natural language representation of formal texts (see e.g. Matuszewski, 1999a; 1999b; 2006), allow to use it as a tool for teaching argumentation theory for those students who are familiar with methods of computer-aided proof checking, applied for example in academic teaching at the faculties of computer science.

However, some applications of Mizar, as discussed in this section, focus exclusively on deductive inference rules and deductive invalidities of reasoning. In order to combine this formal approach with the broader pragmatic account of arguments (as presented in section 2.2), further research on the applications of Mizar is necessary. One of the main goals of such an inquiry would be to analyze, by means of the Mizar language, a set of those tools of argumentation theory which are (at least to some extent) formalizable, and which take into account the context of argument use. Among the tools of argumentation theory which fit to those requirements there are argumentation schemes. The research on representing the main argument schemes in Mizar would be in accord with the attempts at formalizing some argumentation schemes, such as the ad hominem argumentation scheme (Walton, 2010). The fact that some argumentation schemes are generalized rules of inference (Prakken, 2010; see also Bex & Reed, 2011, this issue) constitutes an additional justification for such an inquiry, because, as discussed examples show, representing inference rules is also possible in Mizar. Hence, the task for further inquiry would consist in expressing in the Mizar language those schemes which have the form of generalized inference rules.

## References

Ajdukiewicz, K. (1974). *Pragmatic Logic* (O. Wojtasiewicz, Trans.). Dordrecht/Boston/Warsaw: D. Reidel Publishing Company & PWN – Polish Scientific Publishers. (Original work published 1965). [English translation of *Logika pragmatyczna*].

Amgoud, L., & Dupin de Saint-Cyr, F. (2011). The quality of persuasion dialogs. *Studies in Logic, Grammar and Rhetoric 23(36)*, 69–98.

Bex, F., & Budzyńska, K. (2010). Argument and explanation as contexts of reasoning. In *2010 Workshop on Computational Models of Natural Argument (CMNA).*

Bex, F., & Reed C. (2011). Schemes of inference, conflict, and preference in a computational model of argument. *Studies in Logic, Grammar and Rhetoric 23(36)*, 39–58.

Bryniarski, E., Bonikowski, Z., Waldmajer, J., & Wybraniec-Skardowska, U. (2011). Realistic Premises of Epistemic Argumentation for Dynamic Epistemic Logics. *Studies in Logic, Grammar and Rhetoric 23(36)*, 173–187.

Budzyńska, K. (2004). Argumentation from semantic and pragmatic perspective. *Studies in Logic, Grammar and Rhetoric, 7 (20)*, 127–147.

Budzyńska, K., & Dębowska, K. (2010). Dialogues with conflict resolution: goals and effects. In *SemDial 2010: 14th Workshop on the Semantics and Pragmatics of Dialogue*, Poznań, 16–18 June.

Budzyńska, K., Kacprzak, M., & Rembelski, P. (2009). Perseus. Software for analyzing persuasion process. *Fundamenta Informaticae, 93(1–3)*, 65–79, IOS Press.

Budzyńska, K., & Kacprzak, M. (2011). Model checking of persuasion in multi-agent systems. *Studies in Logic, Grammar and Rhetoric 23(36)*, 99-122.

Calculemus Home Page, www.calculemus.org.

Carbogim, D. V., Robertson, D., & Lee, J. (2000). Argument-based applications to knowledge engineering. *Knowledge Engineering Review, 15 (2)*, 119–149.

Chesñevar, C. I., & Maguitman, A.G. (2000). Logical Models of Argument. *ACM Computing Surveys, 32 (4)*, 337–383.

Chesñevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, & G., Willmott, S. (2006). Towards an Argument Interchange Format. *Knowledge Engineering Review, 21 (4)*, 293–316.

Dębowska, K., Łoziński, P., & Reed, C. (2009). Building Bridges between Everyday Argument and Formal Representations of Reasoning. *Studies in Logic, Grammar and Rhetoric, 16 (29)*, 95–135.

Dung, P.M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence, 77 (2)*, 321–357.

Falkenberg, G. (1996). A hundred years of analytic philosophy. Comments on Peter Simons' study "On Philosophy and Logic in Central Europe". *Mathesis Universalis 1.* Retrieved from http://www.calculemus.org/MathUniversalis/1/index.html.

Gabbay, D., & Ohlbach, H.J. (1996). *Practical Reasoning, International Conference on Formal and Applied Practical Reasoning, FAPR '96*, Bonn, Germany, June 3–7, 1996, Proceedings, Springer.

Gibson, A., Rowe, G., & Reed, C. (2007). A Computational Approach to Identifying Formal Fallacy. In *Working Notes of the 7th Workshop on Computational Models of Natural Argument (CMNA 2007)*, (pp. 26–30), Hyderabad.

Gordon, T.F. (1995). *Pleadings Game: An Artificial Intelligence Model of Procedural Justice.* Kluwer Academic Publishers.

Grabowski, A., Korniłowicz A., & Naumowicz A. (2010). Mizar in a Nutshell. *Journal of Formalized Reasoning*, vol. *3*, no. *2*, 153–245.

Jacquette, D. (2007). Deductivism and the informal fallacies. *Argumentation, 21*, 335–347.

Jacquette, D. (2009). Deductivism in formal and informal logic. *Studies in Logic, Grammar and Rhetoric, 16 (29)*, 189–216.

Jadacki, J. J. (2006). The Lvov-Warsaw School and its influence on Polish philosophy of the second half of the 20th century. In J. Jadacki, & J. Paśniczek (Eds.), *The Lvov-Warsaw School – The New Generation* (pp. 39–83), Amsterdam-New York: Rodopi.

Jadacki, J. J. (2009). *Polish Analytical Philosophy. Studies on Its Heritage.* Warsaw: Wydawnictwo Naukowe Semper.

Jaśkowski, S. (1934). On the Rules of Supposition in Formal Logic. In the Series *Studia Logica: Wydawnictwo Poświęcone Logice i jej Historii* (publications on logic and its history), ed. by Jan Łukasiewicz, no. 1, Warszawa 1934, published by the Philosophical Seminary of the Faculty of Mathematics and Natural Sciences, Warsaw University. Reprinted in (McCall 1967).

Kacprzak, M., Lomuscio, A., Niewiadomski, A., Penczek, W., Raimondi, F. & Szreter, M. (2006). Comparing BDD and SAT based techniques for model checking Chaum's Dining Cryptographers Protocol. Fundamenta Informaticae, vol. *72*, nr *1–3*, 215–234.

Kirschner, P.A., Buckingham Shum, S. J., & Carr, C. S. (2003) (Eds.). *Visualizing Argumentation*. London: Springer Verlag.

Krause, P., Ambler, S., Elvang-Gøransson, M., & Fox, J. (1995). A Logic of Argumentation for Reasoning under Uncertainty. *Computational Intelligence, 11 (1)*, 113–131.

Koszowy, M. (2010). Pragmatic logic and the study of argumentation. *Studies in Logic, Grammar and Rhetoric, 22 (35)*, 29–45.

Kwiatkowski, T. (1993). Classifications of reasonings in contemporary Polish philosophy. In F. Coniglione, R. Poli & J. Woleński (Eds.), *Polish Scientific Philosophy. The Lvov-Warsaw School* (pp. 117–167), Amsterdam: Rodopi.

Lewiński, M. (2011). Dialectical trade-offs in the design of protocols for computer-mediated deliberation. *Studies in Logic, Grammar and Rhetoric 23 (36)*, 225–241.

Lin, F., & Shoham, Y. (1989). Argument systems: a uniform basis for nonmonotonic reasoning. In *Proceedings of the 1st International Conference on Knowledge Representation and Reasoning (KRR'89)* (pp. 245–255), Morgan Kaufmann.

Łoziński, P. (2011). An algorithm for incremental argumentation analysis in Carneades. *Studies in Logic, Grammar and Rhetoric 23 (36)*, 155–171.

Łuszczewska-Romahnowa, S. (1962). Pewne pojęcie poprawnej inferencji i pragmatyczne pojęcie wynikania (A certain concept of valid inference and the pragmatic concept of following). *Studia Logica, 13*, 203–208.

Marciszewski, W. (1991). Foundations of the art of argument. *Logic Group Bulletin, 1*, Warsaw Scientific Society, 45–49.

Marciszewski, W. (1994). A Jaśkowski-style system of computer-assisted reasoning. In J. Woleński (Ed.), *Philosophical Logic in Poland*, Dordrecht: Kluwer. Revised version: A system of suppositional logic as embodied in the proof checker Mizar MSE. *Mathesis Universalis, 3.* Retrieved from http://www.calculemus.org/MathUniversalis/3/index.html.

Marciszewski, W. (1997). Leibniz's Idea of Automated Reasoning Compared with Modern AI. *Studies in Logic, Grammar and Rhetoric, 1 (14).*

Marciszewski, W. (2003). On mechanization of reasoning, decidability of logic, and uncomputable numbers. *Studies in Logic, Grammar and Rhetoric, 6 (19)*, 79–93.

Marciszewski, W. (2006a). The Gödelian speed-up and other strategies to address decidability and tractability. *Studies in Logic, Grammar and Rhetoric, 9 (22)*, 9–29.

Marciszewski, W. (2006b). Undecidability and intractability in social sciences. *Studies in Logic, Grammar and Rhetoric, 9 (22)*, 143–174.

Marciszewski, W. (2007). Computational dynamics of complex systems: a new way of doing science. *Studies in Logic, Grammar and Rhetoric, 11 (24)*, 107–123.

Marciszewski, W., & Murawski R. (1995). *Mechanization of Reasoning in a Historical Perspective*, Amsterdam etc: Rodopi.

Matuszewski, R. (1999a). On natural language presentation of formal mathematical texts. *Studies in Logic, Grammar and Rhetoric, 3 (16)*, 71–80.

Matuszewski, R. (1999b). A structural approach to human oriented presentation of formal proofs. *Studies in Logic, Grammar and Rhetoric, 3 (16)*, 81–93.

Matuszewski, R. (2006). On computer-assisted approach to formalized reasoning. *Studies in Logic, Grammar and Rhetoric, 9 (22)*, 61–71.

Matuszewski, R., & Rudnicki P. (2005). MIZAR: the first 30 years. *Mechanized Mathematics and Its Applications, 4 (1)*, 3–24.

McBurney, P., & Parsons, S. (2002). Dialogue Games in Multi-Agent Systems. *Informal Logic, 22 (3)*, 257–274. McCall, S. (Ed.) (1967). *Polish Logic in 1920-1939.* Oxford: Clarendon Press.

Mizar Home Page, http://mizar.org/.

Pąk, K. (2010). The algorithms for improving and reorganizing natural deduction proofs. *Studies in Logic, Grammar and Rhetoric, 22 (35)*, 95–112.

Prakken, H. (2010). On the nature of argument schemes. In C. Reed and C. Tindale (Eds.), *Dialectics, Dialogue and Argumentation. An Exa-*

*mination of Douglas Walton's Theories of Reasoning and Argument* (pp. 167–185), London: College Publications.

Prakken, H., & Vreeswijk, G. (2002). Logics for Defeasible Argumentation. In D. Gabbay, & F. Guenthner (Eds.), *Handbook of Philosophical Logic* (pp. 218–319), vol. 4, 2nd Edition, Dordrecht etc.: Kluwer Academic Publishers.

Reed, C., & Grasso, F. (2007). Recent Advances in Computational Models of Argument. *International Journal of Intelligent Systems, 22 (1)*, 1–15.

Reed, C., & Norman, T. J. (2003) (Eds.). *Argumentation Machines.* Dordrecht: Kluwer.

Reed, C., & Norman, T. J. (2003b). A roadmap of research in argument and computation. In C. Reed & T. J. Norman (Eds.), *Argumentation Machines* (pp. 1–12), Dordrecht: Kluwer.

Scholz, H. (1930). *Abriss der Geschichte der Logik.* Berlin: Junker and Dünnhaupt.

Stefanowicz, K. (2011). New Tools for Social Dialogue on the Internet. Opportunities and Threats for New Social Sphere. *Studies in Logic, Grammar and Rhetoric 23 (36)*, 243–258.

Trybulec, A. (1993). Some features of the Mizar Language. In *Proceedings of the ESPRIT Workshop*, Torino. Retrieved from http://mizar.uwb.edu.pl/project/trybulec93.ps.

Trzęsicki, K. (2007). Polish logicians' contribution to the world's informatics. *Studies in Logic, Grammar and Rhetoric, 11 (24)*, 15–34.

Trzęsicki, K. (2010). Philosophy of Mathematics and Computer Science. *Studies in Logic, Grammar and Rhetoric, 22 (35)*, 47–80.

Trzęsicki, K. (2011). Arguments and their classification. *Studies in Logic, Grammar and Rhetoric 23 (36)*, 59–67.

Visser, J., Bex, F., Reed, C., & Garssen, B. (2011). Correspondence between the Pragma-Dialectical Discussion Model and the Argument Interchange Format. *Studies in Logic, Grammar and Rhetoric 23 (36)*, 189–224.

Walton, D. (2010). Formalization of the Ad Hominem argumentation scheme. *Journal of Applied Logic, 8*, 1–21.

Walton, D. (2011). How to refute an argument using artificial intelligence. *Studies in Logic, Grammar and Rhetoric 23 (36)*, 123–154.

Walton, D., Reed, C. & Macagno, F. (2008). *Argumentation Schemes.* Cambridge etc.: Cambridge University Press.

Wiedijk F. (2011). Writing a Mizar article in nine easy steps. Retrieved from http://www.cs.ru.nl/ freek/mizar/mizman.pdf.

Woleński, J. (1988). Klasyfikacje rozumowań (Classifications of reasoning). *Edukacja Filozoficzna, 5*, 23–51.

Woleński, J. (1989). *Logic and Philosophy in the Lvov-Warsaw School.* Dordrecht/Boston/Lancaster: D. Reidel.

Woleński J. (1995). Mathematical logic in Poland 1900-1939: people, circles, institutions, ideas. *Modern Logic, 5*, 363–405.

Woleński, J. (2008). Argumentacja i perswazja w filozofii (Argumentation and persuasion in philosophy). In A. Brożek & J. J. Jadacki (Eds.), *VIII Polski Zjazd Filozoficzny* (Proceedings of the VIII-th Polish Philosophical Congress) (pp. 98–117), Warsaw: Wydawnictwo Naukowe Semper.

Wybraniec-Skardowska U. (2009). Polish Logic. Some Lines from a Personal Perspective. *Publications of Institute for Logic, Language and Computation*, Amsterdam. Retrieved from www.illc.uva.nl/Publications/ResearchReports/PP-2009-31.text.pdf.

Chris Reed
Argumentation Research Group
School of Computing
University of Dundee
Dundee DD1 4HN Scotland, UK
chris@computing.dundee.ac.uk

Marcin Koszowy
Chair of Logic, Informatics and Philosophy of Science
University of Białystok, Poland
koszowy@uwb.edu.pl

**Floris Bex and Chris Reed**
University of Dundee

# SCHEMES OF INFERENCE, CONFLICT, AND PREFERENCE IN A COMPUTATIONAL MODEL OF ARGUMENT

**Abstract**: Argumentation demands that various non-deductive patterns of reasoning are accounted for from a strong theoretical foundation. The theory of argumentation schemes has provided such a theoretical foundation, and has led to a significant programme of research not only in epistemological and metaphysical philosophy but also in knowledge representation and multi-agent systems in artificial intelligence. More recently, work on computational models of argument has posited that not only inference, but also conflict, might be governed by more sophisticated relationships that just propositional negation. And finally, work on developing a standard computational ontology for handling argument has suggested that preference too demands such schematization. This paper shows how schematic templates can be designed to capture these stereotypical patterns of inferring, conflicting and preferring, and furthermore, demonstrates the strong representational and constitutive similarities between these apparently very different phenomena.

**Keywords**: argumentation schemes, Argument Interchange Format, inference, conflict, preference

## 1. Introduction

The theory of argument is a rich, interdisciplinary area with insights from diverse disciplines such as philosophy, law, psychology, communication studies and artificial intelligence. This paper explores the ways in which parts of arguments can be connected together.

Recent research in philosophy has shown that the broad range of ways in which inference is performed in natural texts can be understood by taxonomizing and classifying ‘argumentation schemes’, which capture stereotypical patterns of reasoning (Walton 1996; Walton et al. 2008). These argumentation schemes have been demonstrated to be not only powerful tools for scholarly investigation of argument, but also of practical use both in pedagogy and in computational settings (Reed and Walton, 2005). In addition to inference, however, argument makes fundamental use of two further types

ISBN 978–83–7431–305–6 ISSN 0860-150X 

of relation: conflict and preference. Conflict acts as a driver for argumentative discourse, and for many authors is a defining feature of such linguistic behaviour. Preference, in turn, is the key to resolving conflict, particularly where (as is very common) the conflict is rooted not just in propositional disagreement, but in mismatches in values.

This paper argues for an approach that tackles inference, conflict and preference as genera of a more abstract class of schematic relationships. This allows the three types of relationship to be treated in more or less the same way, meaning that the logical and semantic machinery required for handling them is greatly simplified. The context of the work is a method for representing argument structures which aims simultaneously to provide a language that is rich enough to talk about the enormous variety of naturally occurring argument, whilst at the same time enforcing a level of specificity and clarity that allows for computational interpretation. This context is the Argument Interchange Format, AIF, which serves as an interlingua between various software tools and systems in the burgeoning community in computational models of argument (Chesñevar et al., 2006).

The rest of this paper is organised as follows. In section 2 we provide a brief and very general introduction to the most important concepts in argumentation theory. Section 3 introduces the language of the AIF; 3.1 discusses the basic concepts and 3.2 concentrates on the various schematic relations. Section 4–6 discuss inference, conflict and preference schemes, respectively. These sections start with a short introduction to the representation of inference, conflict and preference in models of computational argument. After this, the modelling of these concepts in the language of the AIF is presented. Section 7 concludes the paper.

## 2. Argumentation

In an argument, a defeasible inference leads from premises to a conclusion; associated with a defeasible inference is a generalization, usually in a conditional form, which justifies or *warrants* the inference link between premises and conclusion. Generalizations are generalized statements about how we think the world around us works; they can express generally accepted patterns (e.g. ("If a witness testifies that '*P*' then *P* is the case") or they can be more case-specific (e.g. "Chris is usually at work before 8 o' clock"). Very often, generalizations are left implicit in natural argument, but explicitly expressing the generalization can help in determining the relevance and force of the inference. Take, as a simple example the argument for the conclusion that Harry was in Dundee based on Bob's testimony, visualised as a diagram in the style of Toulmin (2003):

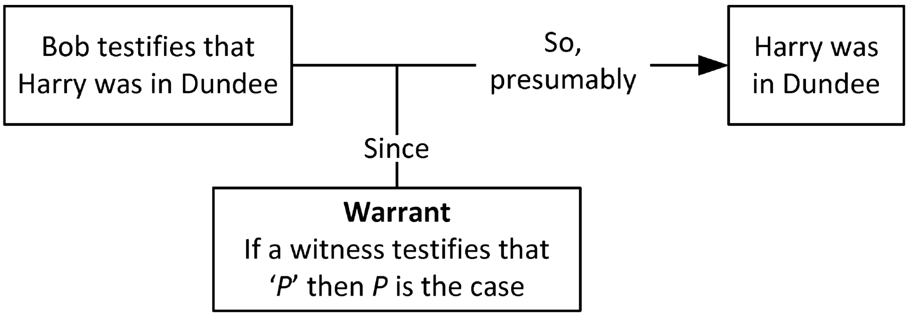



**Figure 1: a Toulmin-style argument for the claim that Harry was in Dundee**

Generalizations that occur often in natural argument have been studied in the form of *argumentation schemes* (Walton et al. 2008), stereotypical patterns of reasoning.[1] As an example, take the scheme for "argument from appeal to witness testimony", which is similar to the above generalization (adapted from Bex et al. 2003):

Witness *W* asserts that *P* is true (false).
*Therefore*, *P* may plausibly be taken to be true (false).

Associated with each argumentation scheme are critical questions that point to standard sources of doubt. Standard sources of doubt with regards to witness testimony are, for example, the witness' bias, whether he is lying or whether he correctly remembers what he observed; critical questions for the argumentation scheme are hence, for example, 'Is the witness biased?' or 'Is there a chance that the witness misremembers?'.

Most everyday arguments are defeasible, in that new information can cast doubt on information previously taken to be true. For example, 'witness Peter testifies that Harry was in Amsterdam' is a reason for the fact that Harry was in Amsterdam. This provides a counterargument to the original conclusion that Harry was in Dundee. In addition to attacking conclusions (called rebuttal in the literature, see Pollock 1994, Prakken 2010), we may also attack the defeasible inference (this type of attack is often called *under-cutting*). Recall that the generalizations or schemes that justify inferences

[1] As Prakken (2010) has shown, argumentation schemes are often (but not always) generalized rules of inference.

express a stereotypical pattern of everyday reasoning: normally we expect that people bear witness only to events they actually observed. However, it is not unthinkable that in any particular case the witness misremembers or lies (cf. the critical questions for the argument from witness testimony). In such a case we are dealing with an exception to the general rule. Such an exception does not deny the premise or conclusion of the argument but attacks the inference from premise to conclusion: if, in the example, the witness is lying, this does not mean that Harry was *not* in Dundee; it just shows that this particular witness testimony is not a good reason for believing this conclusion.

Undercutting and rebutting are just ways to express conflict in argumentation. Conflict is just as important as inference in argumentation: the dialectical process is essentially a process of argument and counterargument. However, while some of the mathematical properties of conflict have been extensively studied,[2] (context-)specific types of conflict have not received much attention in (computational) argumentation theory.[3]

Inference and conflict allow us to build arguments and provide counterarguments. In many contexts, a choice then needs to be made as to which of the arguments one decides to believe or, in other words, which of the arguments is *preferred*. This preference is naturally tied to the applicable "rules" of the discussion (e.g. a judge or jury cannot decide for an argument on inadmissible evidence, even if she prefers this argument). In general, however, this preference is intimately tied to the beliefs of the person doing the evaluation. For example, we can only accept that Harry was in Dundee if we believe that Bob (who stated Harry was in Dundee) is a more trustworthy witness than Peter (who stated that Harry was in Amsterdam). Formal models of argumentation have long enjoyed rich, mature models of preference and priority. Bench-Capon (2003), for example, has shown how one's values might influence the choice of beliefs and Modgil (2009) has extended Dung's (1995) abstract argumentation frameworks with reasoning about preferences. However, as with conflict, more context-specific patterns of preference (outside the value orderings of Bench-Capon) have not been widely examined in (computational) argumentation theory.

[2] See the large body of work on argumentation theoretic semantics in the style of (Dung 1995), e.g. (Caminada 2006, Dunne 2009).

[3] However, some argumentation schemes, such as the scheme for ad hominem arguments, seem to have more to do with conflict rather than inference.

## 3. The Argument Interchange Format

Argumentation is a large and diverse field stretching from analytical philosophy to communication theory and social psychology. The computational investigation of the space has multiplied that spectrum by a diversity of its own in semantics, logics and inferential systems. One of the problems associated with the diversity and productivity of the field, however, is fragmentation. With many researchers from various backgrounds focusing on different aspects of argumentation, it is increasingly difficult to reintegrate results into a coherent whole; for the plethora of methods, processes and tools for argumentation, there are just as many individual languages for argumentation, ranging from logical to visual to natural language. This fragmentation makes it difficult to present new ideas which can be adapted across the board and difficult for new research to build upon old. To tackle this problem, the computational argument community has initiated an effort aimed at building a common ontology for argument: the Argument Interchange Format (AIF).

The AIF is a communal project which aims to consolidate some of the defining work on (computational) argumentation (Chesñevar et al. 2006). The AIF project aims to present a common vision and consensus on the concepts and technologies in the field, thus promoting research and development of new argumentation tools and techniques. A main aspiration of the AIF is to facilitate data interchange among various tools and methods for argument analysis, manipulation and visualization. To this end, the AIF project aims to develop a commonly agreed-upon *core ontology* that specifies the basic concepts used to express argumentative information and relations. The purpose of this ontology is not to replace other languages for expressing argument but rather to serve as an abstract *interlingua* that acts as the centrepiece to multiple individual languages. These argument languages may be, for example, logical languages (e.g. ASPIC, see Prakken 2010), visual diagramming languages (e.g. Araucaria, see Reed and Rowe 2004) or natural languages (e.g. pragma-dialectics, see van Eemeren and Grootendorst 2004). The idea is that an interlingua drastically reduces the number of translation functions that are needed for the different argumentation languages to engage with each other; only translation functions to the AIF core ontology have to be defined (i.e., $n$ instead of $n^2$ functions for $n$ argumentation languages).

### 3.1. The AIF core ontology

In the AIF ontology, arguments and their mutual relations are described by conceiving of them as a an *argument graph*. The ontology falls into two natural halves: the Upper Ontology and the Forms Ontology. The Upper Ontology, introduced in (Chesñevar et al. 2006), describes the language of different types of nodes and edges with which argument graphs can be built. The Forms Ontology, introduced by (Rahwan et al. 2007), allows for the conceptual definition of the elements of the graphs, that is, it describes the argumentative concepts instantiated by the specific nodes in a graph. Figure 2 visually renders part of the ontological structure of the AIF; the explanation of the different elements is below Figure 2. Note that here, only a part of the ontology is shown; as we will show in this paper, for example, conflict schemes also have descriptions of the elements that are in conflict. For readability, however, only the elements connected to the defeasible and deductive inference schemes are shown.

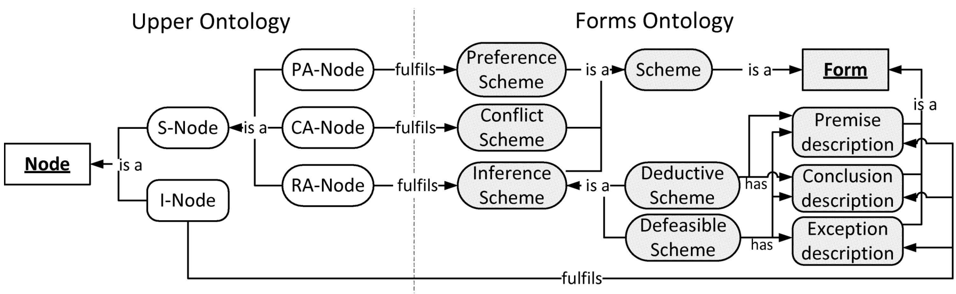



**Figure 2: The AIF core ontology**

The Upper Ontology places at its core a distinction between *information*, such as propositions and sentences, and *schemes*, general patterns of reasoning such as inference or conflict, which are used to relate pieces of information (I-nodes) to each other. Accordingly, there are two types of nodes for building argument graphs: information nodes, *I-nodes*, and scheme nodes, *S-nodes*. Individual nodes can have various attributes (e.g. "creator", "date"). In a graph, I-nodes can only be connected to other I-nodes via S-nodes, that is, there must be a specific scheme application that expresses the rationale behind the relation between I-nodes. In the basic AIF ontology, scheme nodes can be rule application nodes (*RA-nodes*), which denote specific inference relations, conflict application nodes (*CA-nodes*), which denote specific conflict relations, and preference application nodes (*PA-nodes*), which denote specific preference relations. Different S-nodes can be connected to each other; for example, we can express that two preference applications are in conflict with each other (e.g., $x > y$ and $y > x$) by connecting the two PA-nodes through a CA-node.

### 3.2. Scheme application in the AIF

The Upper Ontology defines the basic building blocks of argument-graphs (in a sense, it defines the "syntax" for our abstract language). In contrast, the Forms Ontology defines what these individual nodes mean in argumentative terms It defines the forms of the schemes that are used in reasoning, that is, the inference schemes, conflict schemes and preference schemes. Informally, inference schemes are criteria for inferring (deductively, inductively or presumptively), conflict schemes are criteria (declarative specifications) defining conflict (which may be logical or non-logical) and preference schemes express (possibly abstract) criteria of preference. These main scheme types can be further classified. For example, inference schemes can be deductive or defeasible. Defeasible inference schemes can be further subdivided into more specific argumentation schemes, such as Expert Opinion, Practical Reasoning and so on (see, for example, Walton et al. 2008).[4] Accordingly, the AIF ontology has a Schemes Ontology, which is a sub-ontology of the Forms Ontology. This Schemes Ontology contains specific inference schemes and may vary from very simple (containing only the basic deductive and defeasible schemes) to extensive (containing a large number of specific deductive and defeasible argumentation schemes).

As can be seen in Figure 2, the Forms Ontology and the Upper Ontology are intimately connected because a specific applications of schemes (denoted by RA-, CA- and PA-nodes) are instantiations of general (inference-, conflict- and preference-) schemes; in other words, the S-nodes *fulfil* the schemes expressed in the Forms Ontology. Like argument-graphs from the abstract language of the AIF, schemes can also be translated into a more concrete language; for example, Rahwan et al. (2010) define schemes as combinations of classes of statements in Description Logic, with object level arguments then being instances of those classes. In this paper, like in (Rahwan et al. 2007), we will represent the Forms Ontology and the schemes contained in it as graphs.[5]

RA-, CA- and PA-nodes capture the passage or the process of inferring, conflicting and preferring, respectively, whilst the inference schemes, conflict schemes and preference schemes embody the general principles expressing how it is that A is inferable to B, A is contrastable to B ('conflictable' is too

---

[4] It is important to note that the AIF ontology does not (and should not) legislate as to which schemes or forms are the correct ones; different schemes are each plausible according to particular theoretical assumptions.

[5] Note that these graphs simply express concepts (i.e. Forms) and the ontological relations between them; they are *not* AIF argument graphs, which exist at the object level.

cumbersome a term), and A is preferable to B, respectively. RA-nodes thus correspond mostly closely to what a traditional system of formal logic would regard as entailment, i.e. where $\varphi$ is a premise for an RA to a conclusion $\psi$, the RA corresponds to $\varphi \vdash \psi$. In contrast, conditionals such as $\varphi \rightarrow \psi$ are available as I-nodes – in instances of (defeasible) modus ponens, for example. Of course it is possible to formulate – in natural language – a proposition corresponding to the fact that $\varphi \vdash \psi$, so in principle we can also represent such a proposition as an I-node. But this I-node can be handled as a special type of 'calculated property' (Reed 2010): a propositional result of running some (arbitrary) process over an AIF structure. This proposition could use the RA itself as a basis for establishing the entailment proposition, but as Bex et al. (2010) have argued, this exact connection between an AIF graph and the properties calculated on the basis of the graph cannot be captured in the core AIF ontology itself (and nor should they be, for otherwise the AIF would swell to some general purpose programming language).

This contrast between propositions expressing implicative relationships and propositions expressing entailment relationships is important because for inference, we have strong intuitions and mature theory to guide the way in which the AIF should handle them. For conflict and preference, we need to develop strong analogues to inferential components. If we say then that $\varphi \rightarrow \psi$ expresses that $\psi$ is inferable from $\varphi$, we might similarly say that $\varphi \succ \psi$ expresses that $\varphi$ is preferable to $\psi$, and $\varphi {-\!\times} \psi$ expresses that $\varphi$ is contrastable with $\psi$. These could all be captured by I-nodes and could all serve as foundations for RA-, CA- and PA-nodes respectively. In contrast, $\varphi \vdash \psi$ captures that $\psi$ is (in fact) inferred from $\varphi$, and so similarly we might say that $\varphi \mid\!\succ \psi$ corresponds to the fact that $\varphi$ is (in fact) preferred to $\psi$, and that $\varphi \mid\!\times \psi$ corresponds to the fact that $\varphi$ does (in fact) conflict with $\psi$. Again, these can all be captured by I-nodes, but their connection to RA-, CA- and PA-nodes is tenuous and is governed by the process which determines these calculated properties, and not by the AIF per se. These strong analogies between the three schematic classes are very useful in developing accounts of scheme usage through the AIF as a whole.

## 4. Schemes of Inference

One of the main issues of argumentation discussed in section 2 concerns the generalizations warranting the defeasible inferences. In a logic, conditional generalizations of the form 'if $\varphi$ then $\psi$' (or'$\varphi$ therefore $\psi$') can be modelled both as an object-level rule ($\varphi$ *implies* $\psi$, formally represented as

$\varphi \rightarrow \psi$) or as a metalinguistic rule of inference ($\varphi$ *entails* $\psi$, formally represented as $\varphi \vdash \psi$). The various argumentation logics (see Chesñevar et al. 2000, Prakken and Vreeswijk 2002 for overviews) have different stances on how these generalizations should be modelled. For example, Prakken (2010) and Pollock (1994) model them as rules of inference, whereas (Bondarenko et al. 1997) and Verheij (2003) model them as (defeasible) implications in the object language and have a (defeasible) modus ponens inference rule for reasoning with these implications. The most important difference between the two ways of modelling them is that conditionals in the object language can be reasoned about in a natural way; they can, for example, be denied by arguing that "$\neg(\varphi \rightarrow \psi)$" or they can serve as the conclusions of arguments. A problem for the argumentation logics that model generalizations as inferences is that it is often unclear how statements like $\varphi \vdash \psi$ can be rendered in the object language and what, *in the object language*, their relation to $\varphi \rightarrow \psi$ is. In the AIF ontology, the fact that in the AIF $\varphi \vdash \psi$ is represented by its own RA-node in the object language and $\varphi \rightarrow \psi$ is represented by its own I-node in the object language disambiguates this relationship between the two.

Now, as was argued above conditional generalizations can be modelled either in the object language (as an I-node) or in the metalanguage, as a Scheme in the Forms Ontology. Figure 3a models the conditional expressing the generalization as a premise (I-node) and connects this premise together with another premise (I-node) representing the antecedent of the conditional to the conclusion by way of a (defeasible) modus ponens inference (RA-node).

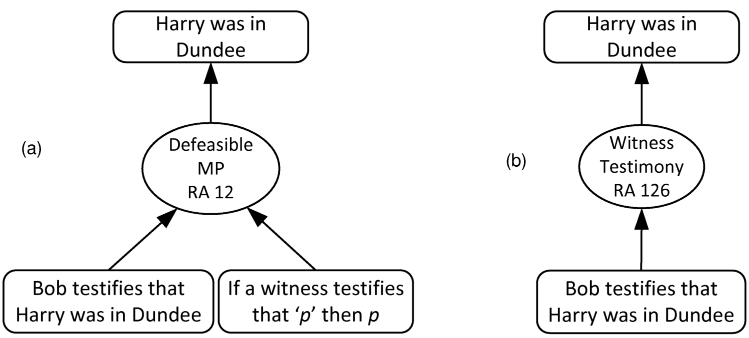



**Figure 3: Two ways of modelling defeasible inference**

The inference rule that is applied is explicitly shown in the AIF structure. In the case of the argument in Figure 3a, the generalization justifying the inference ("If a witness testifies that '$P$' then $P$") is made explicit as an

I-node and can be questioned. However, no further information about the generalization is provided; if an arguer or an analyst wishes to critique the inference step (e.g. by undercutting it), it remains for them to introduce sufficient contextual knowledge to form an attack. One of the advantages of the scheme-based approach advocated by, among others, Walton et al. (2008) is that it provides a theoretically principled way of structuring this contextual knowledge. So the Argument Scheme from Witness Testimony provides not just a characterisation of the minor premise and conclusion, but also a raft of implicit premises (presumptions) which may be taken to hold, and exceptions, which may be taken not to hold. These presumptions and exceptions are part of the Scheme Ontology. A scheme-based analysis (Figure 3b) shows that the premise and conclusion are connected by this specific type of Witness Scheme inference. The general form of this scheme gives the implicit premises and exceptions are part of the scheme and which can be used to critique the scheme. Figure 4 shows both the abstract Defeasible Modus Ponens (a) and the specific Witness Testimony Scheme (b). Notice that the Witness Testimony scheme shows the (implicit) presumptions and exceptions; exactly how these can be used to attack an argument that uses the scheme will be discussed below in section 5.

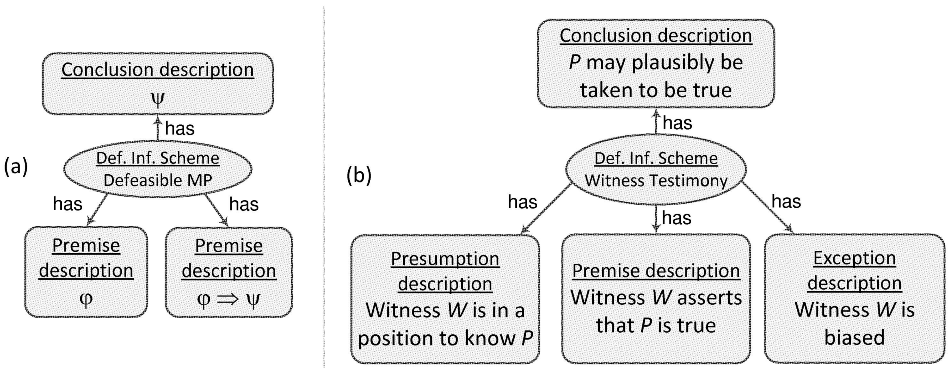



**Figure 4: The Defeasible Modus Ponens and Witness Testimony schemes as represented in the AIF**

So modelling generalizations as conditional premises as in Figure 4a allows for a lot of flexibility, whereas modelling them as schemes as in Figure 4b provides a firm grounding to the rules of inference that are being used in our reasoning. This can be seen in the case of Toulmin's characterisation of backing, a reason for why we should believe the warranting generalization. In at least some of Toulmin's examples, backing serves to justify a general rule, rather than its specific application. This is possible in the case of the argument in Figure 3a: reasons can be given for the conditional premise that expresses the generalization. In the case of the argument in Figure 3b, a backing can only be given if it is explicitly encoded in the Forms Ontology (i.e. if the scheme for Witness Testimony has a Backing description). It is not possible to give a backing in an object-level argument, as this would require Scheme Forms from the Scheme Ontology to be able to stand as the conclusions of arguments.

It is important to note that AIF ontology does not (and should not) legislate as to which analysis in Figure 3 is correct. They are each plausible according to particular theoretical assumptions. Similarly, the AIF ontology does not (and should not) legislate as to which schemes or forms are the correct ones; different schemes are each plausible according to particular theoretical assumptions. Argument analysis needs, like many techniques applicable to naturally occurring language, to be flexible, and to admit of alternative views. AIF's job is to make such alternative analyses clear and unambiguous in a common language.

## 5. Schemes of Conflict

Conflict is a central notion in dialectical argumentation and it can take many forms. For example, two claims may be in conflict because they express opposing points of view or because they were uttered by people from different political parties. In logical models of argument, conflict is often equated with logical conflict, i.e., the contradiction between $\varphi$ and $\neg\varphi$. Some frameworks for formal argumentation (e.g. Bondarenko et al. 1997, Prakken 2010) generalize this to a contrariness relation, where $\varphi$ is in conflict with its contrary $\bar{\varphi}$. Thus, other non-logical conflict relations can be expressed.

An important concept in logical models of argument, which is closely related to conflict, is that of *attack*. Attack expresses that one argument is somehow a counterargument to another.[6] However, conflict is not the same as attack. First, the fact that two propositions are in conflict does not mean they attack each other, as this depends on one's definition of attack. For example, in the ASPIC framework (Prakken 2010) a proposition $\varphi$ only attacks another proposition $\neg\varphi$ if $\neg\varphi$ is not a necessary premise. If this is

[6] Not to be confused with "defeat". Attack and defeat are different concepts: attacking your enemy does not guarantee their defeat, only a successful attack defeats. So attack expresses that one argument is a counterargument to another, whilst defeat says that an argument is a counterargument and is preferred (Garcia and Simari 2004).

the case, $\varphi$ is in conflict with $\neg\varphi$ but it does not attack it. Here, attack is *based* on conflict, it is a calculated property. Second, attack is often defined over arguments, where conflict is usually only defined over propositions and, in some cases, inference applications (Figure 7).

In the AIF ontology, conflict is expressed using *conflict schemes* in the Forms Ontology and applications of these schemes in the object layer, *conflict application* or *CA-nodes*. Conflict schemes are similar to (but certainly not analogous to) inference schemes, in that they are patterns of reasoning which are often used in argumentation. Like inference schemes, conflict schemes may denote abstract, logical patterns (e.g. logical conflict) as well as more concrete patterns of conflict dependent on, for example, legal or linguistic conventions (e.g. a bachelor is not married, a man is not a woman). Like inference schemes, conflict schemes can be strict (no exceptions to the scheme; e.g., $\varphi$ and $\neg\varphi$ are always in conflict) or defeasible (there are exceptions to the scheme; e.g. a man is not a woman unless (s)he is androgynous).

Like inference, conflict is often expressed as a generalization; for instance, "people cannot be in two places at the same time" or "it is impossible for both the Tories and Labour to both be in government". Where generalizations that warrant inference are often rephrased as conditionals of the form "if $\varphi$ then $\psi$", generalizations that express conflict can be rephrased as "$\varphi$ conflicts with $\psi$". These conflict generalizations can be represented as information (I-nodes) in the object layer, or in the layer of the Schemes Ontology, as conflict schemes. Take, for example, the conflict between the British Labour Party and the British Conservative Party ("the Tories") being in government. Generally, the two parties are not in the same government (the last time was during the Second World War). Now, we can make the generalization (Figure 5a), or we can model the conflict generalization a separate conflict scheme (Figure 5b and Figure 6b).

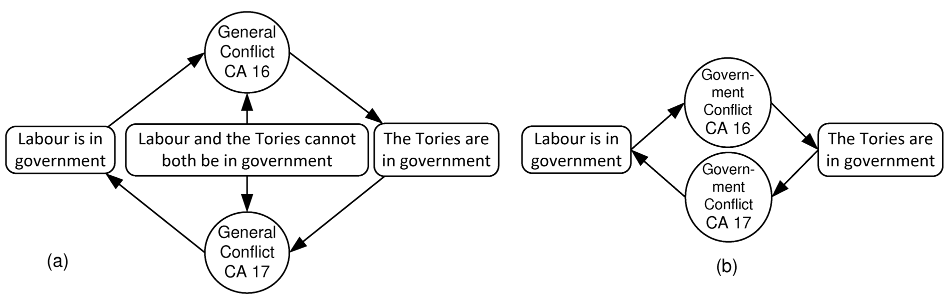



**Figure 5: Two ways of modelling conflict generalizations**

Figure 5 shows an important difference between conflict and inference, namely that often (but not always), conflict is a symmetrical relation, whilst inference is certainly not. That is, if $\varphi$ is in conflict with $\psi$, then $\psi$ is also in conflict with $\varphi$. For inference, this is not the case. One of the reasons for this is that with inference, we can *gain* new information (e.g. we have information that a witness testified that Harry was in Dundee, so we can infer the new information that Harry was in Dundee). Conflict schemes have no such generative function, as they only allow us to represent conflict between existing propositions.

In Figure 5, two conflict schemes are used, a general and a specific one. These two schemes are rendered in Figure 6. The general conflict scheme (Figure 6a) takes a generalization from an I-node and uses this generalization to warrant the application of a conflict. In this sense, it can be likened to the inference scheme for (Defeasible) Modus Ponens (Figure 4a), which warrants the *inference application* with a generalization from an I-node.

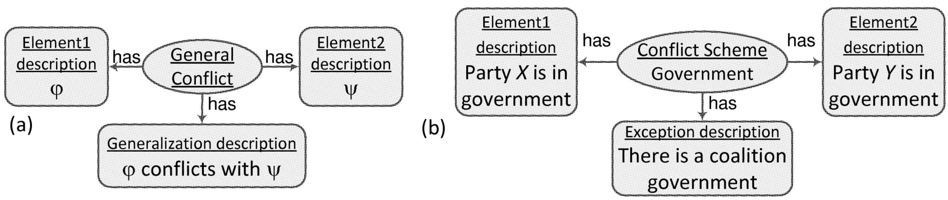



**Figure 6: Conflict schemes in the AIF Forms Ontology**

An advantage of modelling conflict generalizations as I-nodes is that they can be reasoned about. For example, we can give reasons for why, in general, Labour and Tories cannot be in the same government, having the I-node that contains this generalization in Figure 5 as the conclusion of an RA-node. When conflict generalizations are modelled as schemes in the Forms Ontology, it is not possible to provide them with a "backing" in this way. However, representing a conflict generalization as a scheme allows us to specify implicit presumptions and exceptions to the scheme. For instance, an exception to the generalization that Labour and Conservatives are not in the same government is that there is a coalition government, as was the case during the Second World War (the exception basically says that the elements 1 and 2 are in conflict *unless* there is a coalition government of party $X$ and $Y$). Thus, the implicit presumptions and exceptions to conflict relations can be incorporated in a principled way.

Conflict does not just exist between I-nodes. There are cases in which, for example, some information is in conflict with an inference or a preference, or two inferences or preferences are in conflict. Take the example in

Figure 7. Here, the information that Bob is biased conflicts with the application of the Witness Testimony inference scheme. This type of conflict, called undercutting by Pollock (1994), is quite common in argumentation. It allows us to attack the way in which some information has been derived rather than the information itself (that is, we attack $\varphi \vdash \psi$). In the example, knowing that Bob is biased is not a reason for the opposite conclusion, that Harry was *not* in Dundee, but rather it is a reason to believe that we might not be justified in inferring Harry's whereabouts from Bob's testimony. Figure 7b shows the conflict scheme used in the argument. Note how this conflict scheme connects an inference scheme with one of its exceptions.

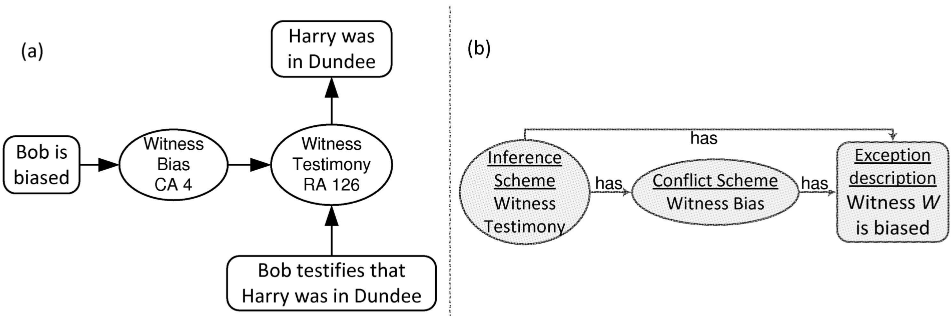



**Figure 7: Conflict between an I-node and an RA-node**

## 6. Preference Schemes in Argumentation

In addition to inference and conflict, the we treat preference as a basic concept of argumentation. Inference and conflict allow us to build arguments and provide counterarguments. In many contexts, a choice then needs to be made as to which of the arguments one decides to believe. Based on the arguments for the prosecution and the defence, does the jury rule the suspect to be guilty or innocent? After a long election campaign, who do we decide to vote for? After comparing the pros and cons, which car (if any) do we buy? The thought that one argument (or set of arguments) is considered better or stronger than another can be expressed using preferences. For example, the jury can argue that the witnesses for the prosecution were more convincing than those for the defence. Which argument we believe may depend on personal preferences; for instance, someone who prefers equality to enterprise and red to blue will generally vote social democrats and buy red cars.

Formal models of argumentation have long enjoyed rich, mature models of preference and priority. For example, (Amgoud and Cayrol 2002, Garcia and Simari 2004) define in, for example, systems of preference-based argumentation, where preferences are used to determine whether an argument that attacks another argument actually defeats the attacked argument. Bench-Capon (2003) further extends this by basing the preferences between arguments on value orderings. Modgil (2009) has proposed Extended Argumentation Frameworks in the style of Dung (1995), where preferences are modelled as attacks on attacks: if an argument A is preferred to another argument B, any attack from B on A is itself attacked. Recently, Prakken (2010) has incorporated preferences in his framework for structured argumentation. Here, the preferences are not between arguments but rather between premises or inference rules. If desired, preferences between arguments can be calculated on the basis of these preferences (Modgil and Prakken 2010).

In line with the now-familiar pattern, the AIF ontology expresses preferences by using *preference schemes* and applications of these schemes in the object layer, *preference application* or *PA-nodes*. Like conflict schemes, preference schemes are similar to inference schemes, again with some important differences, which will be highlighted below. Like inference schemes, preference schemes are patterns of reasoning which are often used in argumentation, which may be abstract, logical patterns as well as more concrete and context-dependent patterns. Preference schemes can also be strict (no exceptions to the scheme) or defeasible (there are exceptions to the scheme).

In argumentation (as in most everyday language use), the preferences themselves can be expressed as generalizations of the form "$\varphi$ is preferred to $\psi$". As with inference and conflict, these generalizations (which can be said to warrant a particular preference) can be explicitly rendered in the object layer, that is, as I-nodes, or they can be modelled as a concrete preference scheme in the Schemes Ontology. So, for example, say that we have a generalization that, in general, government policies that promote equality are preferred over policies that promote enterprise. Figure 8a shows this generalization as an I-node that warrants the application of a general preference scheme and Figure 8b shows this generalization as a specific preference scheme. The preference schemes used in Figure 8 are shown in Figure 9.

Note the similarities with conflict and inference: while modelling the generalization as in Figure 8a allows us to further reason about this generalization, incorporating it as a scheme allows us to provide possible exceptions to this generalization. One example of reasoning about preference generalizations is to base them on one's ideals, one's values (Bench-Capon 2003).

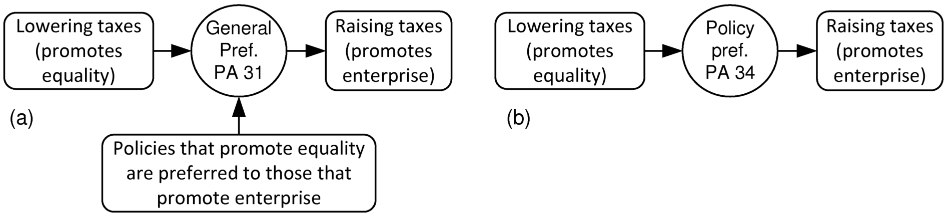



**Figure 8: Two ways of modelling preference generalizations**

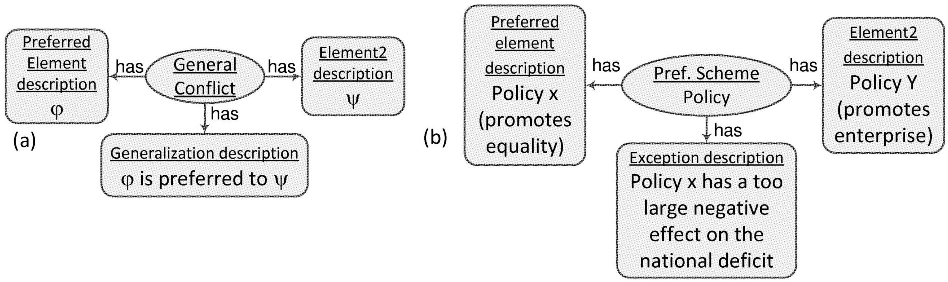



**Figure 9: Preference schemes in the AIF Forms Ontology**

Figure 10 shows how this can be done. The scheme that corresponds to RA74 is not rendered, but will be something along the lines of "if one prefers value A to value B, one should amend one's policies accordingly".

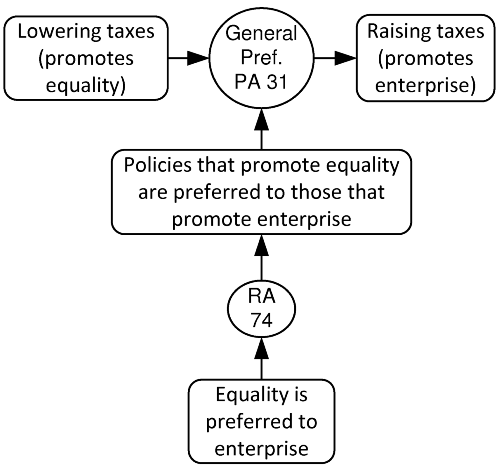



**Figure 10: Two ways of modelling preference generalizations**

As was already discussed in section 4, it is of course possible to provide such a "backing" for the policy preference scheme in the Forms Ontology. As for inference generalizations, rendering them as I-nodes provides flexibility, as the generalization can easily be denied or argued for. Rendering a generalization as a scheme, however, is that it structures contextual knowledge in a principled way. Whilst argumentation schemes for inference are a subject of much study, conflict and preference schemes have not yet been fully developed. Hence, the examples in this paper of such schemes (Figure 6b and Figure 9b) might seem somewhat far-fetched. More intuitive examples the sorts of contextual knowledge preference schemes in the Forms Ontology can express are perhaps the irreflexivity and antisymmetry properties of a particular preference relation. Take, for example, the preference relation $\succ$ as described by (Prakken 2010). A scheme for this relation can be incorporated in the Forms Ontology (Figure 11).

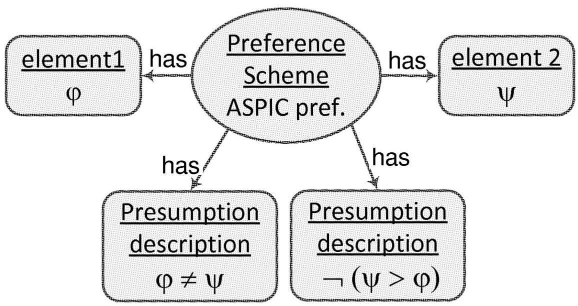



**Figure 11: The ASPIC preference relation as a scheme in the Forms Ontology**

Here, the properties of irreflexivity and antisymmetry have been incorporated into the scheme as implicit presumptions.

## 7. Concluding remarks

In this paper, we have shown how the apparently very different relationships of inference, conflict and preference can be captured analogously in a common language. The approach provides an ontologically parsimonious way of handling a diverse and sophisticated range of argumentation components. Schematising all of these relationships offers particular advantages in terms of explicit characterisation of the constitution of different forms of inference, conflict and preference; spelling out missing or implicit parts (such as assumptions and presumptions), and capturing stereotypical ways of evaluating and critiquing.

We have also shown for the first time how scheme instances can interact with propositional statements that capture expressions of inference, preference and conflict, by virtue of the distinction between, on the one hand, the inferring/preferring/conflicting relation captured by RA/PA/CA-nodes and on the other hand, the inferability/preferability/contrastability captured by I-nodes. Whereas in the current logics for argumentation the distinction between $\varphi \rightarrow \psi$ and $\varphi \vdash \psi$ is fairly well developed, these distinctions are often not explicitly made for preference (i.e. between $\varphi \succ \psi$ and $\varphi \mid\!\succ \psi$) or for conflict (i.e. between $\varphi \!-\!\times \psi$ and $\varphi \mid\!\times \psi$).[7] As an increasing number of research groups and systems start to take advantage of what the AIF has to offer, and thereby, what other teams have already achieved, it becomes vital that a thorough understanding of schematic argument relations and their inter-connections is established, and it is this that the current paper has laid out.

## References

Amgoud, L., and Cayrol, C. (2002). A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence* 34:1, 97–215.

Bex, F. J., Prakken, H., and Reed, C. A. (2010). A formal analysis of the AIF in terms of the ASPIC framework. In P. Baroni, F. Cerutti, M. Giacomin & G. R. Simari (eds.): *Computational Models of Argument. Proceedings of COMMA 2010.* 99–110. Amsterdam: IOS Press.

Bex, F. J., Prakken, H., Reed, C., and Walton, D. N. (2003). Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. *Artificial Intelligence and Law* 11, 125-165.

Bench-Capon, T. J. M. (2003). Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation* 13:3, 429–448.

Bondarenko, A., Dung, P. M., Kowalski, R. A., and Toni, F. (1997). An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence* 93:1–2, 63–101.

[7] One could argue that in Bench-Capon (2003), the value ordering expresses the preferability relation which is used to found the actual preference between propositions (in this case propositions expressing particular government policies). However, as we showed in Figure 10, the value ordering is actually a reason for the preferability between the policies, on which ultimately the actual preference between the policies is founded.

Caminada, M. W. A. (2006). Semi-Stable Semantics. *Computational Models of Argument. Proceedings of COMMA 2006*, 121–130. Amsterdam: IOS Press.

Chesñevar, C. I., Maguitman, A. G., and Loui, R. P. (2000). Logical models of argument. *ACM Computing Surveys (CSUR)* 32:4, 337–383.

Chesñevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., and Willmott, S. (2006). Towards an Argument Interchange Format, *Knowledge Engineering Review* 21 (4), 293–316.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77:2, 321–357.

Dunne, P. E. (2009). The computational complexity of ideal semantics. *Artificial Intelligence*, 173:1559–1591.

Eemeren, F. H. van, and Grootendorst, R. (2004). *A Systematic Theory of Argumentation – The pragma-dialectical approach.* Cambridge: Cambridge University Press.

Freeman, J. B. (1991). *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure.* Berlin: Foris Publications.

García, A. J., and Simari, G. R. (2004). Defeasible Logic Programming: An Argumentative Approach. *Theory and Practice of Logic Programming* 4:1, 95–138.

Modgil, S. (2009). Reasoning About Preferences in Argumentation Frameworks. *Artificial Intelligence* 173:9–10, 901–1040.

Modgil S., and Prakken, H. (2010). Reasoning about preferences in structured extended argumentation frameworks. In P. Baroni, F. Cerutti, M. Giacomin & G. R. Simari (eds.): *Computational Models of Argument. Proceedings of COMMA 2010.* 347–358. Amsterdam: IOS Press.

Pollock, J. L. (1994). Justification and defeat. *Artificial Intelligence* 67 (2), 377–408.

Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation* 1, 93–124.

Prakken, H. (2010). On the nature of argument schemes. In C. A. Reed and C. Tindale (eds.) *Dialectics, Dialogue and Argumentation. An Examination of Douglas Walton's Theories of Reasoning and Argument*, 167–185. London: College Publications.

Prakken, H., and Vreeswijk, G. (2002). Logics for defeasible argumentation. In Goebel, R. and Guenthner, F. (eds.), *Handbook of Philosophical Logic*, 219–318, Dordrecht: Kluwer Academic Publishers.

Rahwan, I., Zablith, F., and Reed, C. (2007). Laying the Foundations for a World Wide Argument Web, *Artificial Intelligence* 171, 897–921.

Reed, C. (2010). Walton's theory of argument and its impact on computational models. In Reed, C. & Tindale, C. W. (eds.) *Dialectics, Dialogue and Argumentation. An Examination of Douglas Walton's Theories of Reasoning and Argument*, 73–84. London: College Publications.

Reed, C. and Rowe, G. (2004). Araucaria: Software for Argument Analysis, Diagramming and Representation, *International Journal of AI Tools* 13 (4), 961–980.

Reed, C. & Walton, D. (2005). Towards a Formal and Implemented Model of Argumentation Schemes in Agent Communication, *Autonomous Agents and Multi-Agent Systems*, 11 (2), pp. 173–188.

Toulmin, S. E. (2003). *The Uses of Argument*, Updated edition (originally published in 1958), Cambridge: Cambridge University Press.

Verheij, B. (2003). DefLog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation* 13:3, 319–346.

Walton, D. N. (1996). *Argumentation Schemes for Presumptive Reasoning*, Mahwah (NJ): Lawrence Erlbaum Associates.

Walton, D., Reed, C., and Macagno, F. (2008). *Argumentation Schemes*, Cambridge University Press.

Floris Bex
Argumentation Research Group
School of Computing
University of Dundee
Dundee DD1 4HN Scotland, UK
florisbex@computing.dundee.ac.uk

Chris Reed
Argumentation Research Group
School of Computing
University of Dundee
Dundee DD1 4HN Scotland, UK
chris@computing.dundee.ac.uk

**Kazimierz Trzęsicki**
University of Białystok

# ARGUMENTS AND THEIR CLASSIFICATION

**Abstract**: The theory of argumentation has ever been the subject of interest of logicians. For some informal logicians the post-Fregean formal logic is not a proper tool of representing natural language and understanding of everyday argumentation. A new stimulus for the theory of argumentation is given by the development of Information and Communication Technologies and their employment in Artificial Intelligence. We will try to define the argument as generally as it is possible to encompass the intuitive notion of the argument involved in the natural language discourse. We develop the concept of the argument as the basis for developing a natural classification of arguments. The argument will be conceived as a pair of nonempty sets of propositions. Propositions will be characterized by their relation to a system of knowledge (a theory or a system of beliefs). The division of arguments conceived as a pair of sets of propositions will be based on the type of relation between the sets and the type of propositions being members of the sets. Finally, we try to clarify how the concept of the argument can assist in developing a classification of arguments.

**Keywords**: argumentation, structure of argumentation, assertion, rejection, suspension

## 1. Argument

In formal and mathematical logic the notion of argument is precisely defined and theoretically elaborated. But this notion does not comprise arguments as they are used in conversation, in metalanguage considerations and in social context as it is the case with juristic arguments. The general notion of the argument is far from clarity.

### 1.1. Propositions in argumentation

Propositions may have different status with respect to ($\mathfrak{B}$) – a particular system of knowledge, a theory or one's system of beliefs. Small Greek letters will be used to denote propositions (simple or compound). Large Greek letters will be used to denote sets of propositions.

Four types of relations between a proposition and $\mathfrak{B}$ can be distinguished. With respect to $\mathfrak{B}$ a proposition can be:

ISBN 978–83–7431–305–6 ISSN 0860-150X

1. asserted: $\mathfrak{B} \vdash \phi$;[1]
2. rejected: $\mathfrak{B} \dashv \phi$;[2]
3. suspended: $\mathfrak{B} \vdash\!\dashv \phi$,
4. neither asserted, nor rejected, nor suspended: $\phi$.

In the case of assertion of a proposition the argumentation for the proposition fulfills the requirements imposed on $\vdash$. A proposition is rejected if there are some reasons that the requirements imposed on $\vdash$ would not be fulfilled. A proposition is suspended with respect to $\vdash$ if there are some arguments for or there are some arguments against the proposition and neither of the arguments is deciding; neither arguments for are satisfactory with respect to $\vdash$, nor arguments against are satisfactory with respect to $\vdash$. A proposition is neither asserted, nor rejected, nor suspended if there are no arguments for or against with respect to $\vdash$.

Any of the first three types of relations are graded. In the natural language discourse, the gradation is described qualitatively. For formal purposes quantitative description would be required. Let „$s$-$\phi$" (**s**igned proposition) denote a proposition of any of the four types of propositions.

We may argue for[3]:

1. assertion,
2. rejection,
3. suspension

any of the $s$-propositions. An $s$-proposition with which the argumentation starts will be called *premiss* (of this argumentation). An $s$-proposition with which the argumentation ends will be called *conclusion* (of this argumentation). Both the notion of the premiss and the conclusion are relative: a proposition that is a premiss (conclusion) of an argumentation may be a conclusion (premiss) of another argumentation. Propositions of any type can be a premiss and can be a conclusion. We may argue, for example, to make higher the degree of assertion of a proposition, or we may argue for the rejection of an asserted proposition. As a premiss a rejected proposition as well as asserted one may be used.

[1] The sign was introduced by G. Frege (1879). According to him, it serves to express a judgment.

[2] The notion of rejection has been introduced to formal logic by J. Łukasiewicz. Formal theory of rejection was developed, e.g. by J. Słupecki and his collaborators, see (1971, 1972).

[3] In the following, if it is clear from the context, $\mathfrak{B}$ will be assumed. Thus, eg. we will write: $\vdash \phi$ instead of $\mathfrak{B} \vdash \phi$.

Consider some examples of different types of *s*-propositions. Suppose that the expression „God exists" is a proposition, i.e. that the sentence has a meaning.

1. There are people who believe in God. Thus these people affirm the existence of God. They assert the proposition: *God exists.*
2. There are people who do not believe in God. Thus these people deny the existence of God. They reject the proposition: *God exits.*
3. There are people who are skeptical about God and there are people – called agnostics – who deny the possibility of finding an answer for the question of existence of God. As well skeptics as agnostics neither affirm nor deny the existence of God. They suspend judgment about whether or not God exists. They suspend the proposition: *God exists.*

Some propositions of the language of mathematics are:

1. asserted: $2 + 2 = 4$;
2. rejected: $2 + 2 = 5$
3. suspended: the Goldbach conjecture – *Every even integer greater than 2 can be expressed as the sum of two primes.*

In physics the proposition:

1. $E = mc^2$ – is asserted,
2. *heavy objects fall faster than lighter ones, in direct proportion to weight* – is rejected,
3. any proposition that with perfect accuracy states position and momentum of a particle is neither asserted nor rejected.[4]

## 1.2. Structure of argumentation

Some premisses as well conclusions are not be written (spoken) directly in a text. These are enthymeme's premisses and conclusions. The set of premisses as well the set of conclusions is conceived as formed by all the written (spoken) and enthymeme's premisses and conclusions. By a text we conceive the sequence of all the sentences that are written (spoken) and that are given implicite (enthymeme). It means a text is a set of sentences indexed by natural numbers.

Argumentation is built out of simple arguments. In text $\mathfrak{T}$ an argument $\langle \Sigma, \Gamma \rangle$, where $\Sigma$ is the set of premisses and $\Gamma$ is the set of conclusions, is a simple argument if and only if:

[4] According to the *Heisenberg uncertainty principle* there is a limit on the accuracy with which certain pairs of physical properties of a particle cannot be simultaneously known.

1. in $\mathfrak{T}$ no element of $\Sigma$ is taken as a conclusion or a premiss of other elements of $\Sigma$,
2. any propositions that in $\mathfrak{T}$ is used as a premiss for $\Gamma$ is an element of $\Sigma$,
3. in $\mathfrak{T}$ no element of $\Gamma$ is taken as a premiss or a conclusion of other elements of $\Gamma$,
4. any propositions that in $\mathfrak{T}$ is used as a conclusion of $\Sigma$ is an element of $\Gamma$.

A simple argument $\langle\Sigma, \Gamma\rangle$ includes:

1. only all premisses of a given set of conclusions $\Gamma$,
2. only all conclusions of a given set of premisses $\Sigma$,
3. in $\mathfrak{T}$ no subset of $\Sigma$ is divided into a set of premisses and a set of conclusions,
4. in $\mathfrak{T}$ no subset of $\Gamma$ is divided into a set of premisses and a set of conclusions.

The simple argument may be described as the largest fragment of $\mathfrak{T}$ that can be divided into a set of premisses and the set of their conclusions and it is the only such division.

There are different relations between the set of premisses, i.e. the set of $s$-propositions with which the argumentation starts and the set of conclusions, i.e. the set of $s$-propositions with which the argumentation ends. We distinguish the following directions:

1. *direction of argumentation*: from the set of premisses to the set of conclusions;
2. *direction of entailment*: $\Sigma$ entails $\Gamma$ ($\Sigma$ logically implies $\Gamma$, or $\Gamma$ is the set of logical consequences of $\Sigma$);
3. *direction of justification*: $\Sigma$ gives evidence that (supports, grounds) $\Gamma$.

The places of premisses and conclusions in the text can be different but the direction of argumentation is determined by the context and special words.

In logic the relation of entailment is defined for propositions (not for $s$-propositions). From $\Sigma$ follows $\Gamma$ if and only if the conjunction of propositions of $\Sigma$ and the negation of disjunction of propositions of $\Gamma$ are inconsistent, i.e. there is no possible situation in that both the propositions would be true. It could be that neither $\Gamma$ follows from $\Sigma$ nor $\Sigma$ follows from $\Gamma$. Thus the relation of entailment is not total.

For Łukasiewicz the division of reasonings into deductive and reductive is more proper than the division into deductive and inductive reasonings.[5]

[5] See (Bocheński 1980, p. 75) or (Bochenski 1992) – in Polish.

The relation of justification holds between sets of $s$-propositions $\Sigma$ and $\Gamma$ if any of $s$-propositions of $\Sigma$ is used in $\mathfrak{T}$ to give evidence for (to support, to ground or is a reason for) assertion, rejection or suspension of at least one of the $s$-propositions of $\Gamma$.

In complex argumentation $s$-propositions are used to support other $s$-propositions. $s$-Propositions that are supported may also be used to support other $s$-propositions.

An elementary unit of argumentation is an argument, which is formed by premisses ($\Sigma$) and their conclusions ($\Gamma$). It means that:

- no $s$-proposition of the set of premisses $\Sigma$ is (in considered argument) a premiss or conclusion of subset of $\Sigma$;
- no $s$-proposition of the set of conclusion $\Gamma$ is a premiss or conclusion of the subset of $\Gamma$.

To describe the structure of argumentation diagrams can be applied.[6] The technique of argument diagramming is used to aid in the identification and analysis of argumentation as well in informal logic, as in legal logic and AI for the representation of knowledge and reasoning. Though the technique is well-established it is still not in an advanced state of development (Reed, Walton & Macagno 2007).

It should be decided which icons will be used to denote:

- direction: argumentation, entailment, giving evidence;
- goal of argumentation: to assert, to reject, or to support suspension;
- type of proposition: asserted, rejected, suspended or neither asserted, nor rejected or suspended.

## 2. Types of reasonings

The question of classifications of reasonings was discussed by Polish logicians, e.g. Łukasiewicz (1915) conceived reasoning as a mental proces of seeking of sentences which entail from a given sentences. In the case of deduction the direction of reasoning is the same as the direction of entailment. In the case of reduction the direction of reasoning is opposite to the direction of entailment. Czeżowski tried to improve Łukasiewicz's classification. Both classifications were criticized by Ajdukiewicz (1965).

[6] Informal logic is mainly conceived and still is developed for educational goals. Thus it is natural to use some didactic improvements which could be helpful in teaching and mastering of reasoning and analyzing skills by students. The traditional square of opposition may be pointed as a device of this type. Frege employed diagrams as the formal language of his *Begriffsschrift*. The "language", due to its intricateness, has not been approved by logicians.

In the case of argumentation we want to show that the methodological requirements of a theory are fulfilled or that the argument is convincing. In the case of reasoning we want to show some reasons for truth-value. Reasoning can be described with the same diagrams that are used to describe argumentation.

In the case of reasoning, if the principle of bivalence is accepted, any argument for the rejection of proposition is equivalent to assertion of the negation of the proposition. From this assumption it follows that in a simple argument we may argue only for assertion of a proposition.[7]

Let $\Sigma$ be a set of premisses, $\Gamma$ – set of conclusions. Let $\vdash \Delta$ be $\{\vdash \phi : \phi \in \Delta\}$. $\Delta \mathrel{-\!\!\!-\!\!\!\dashv} \Lambda$ or $\Lambda \mathrel{\vdash\!\!\!-\!\!\!-} \Delta$ means that the propositions of $\Delta$ are used to give evidence for (to support, to ground) the propositions of $\Lambda$.

There are four combinatorial possibilities:

1. $\vdash \Sigma \mathrel{-\!\!\!-\!\!\!\dashv} \Gamma$,
2. $\vdash \Sigma \mathrel{\vdash\!\!\!-\!\!\!-} \Gamma$,
3. $\Sigma \mathrel{-\!\!\!-\!\!\!\dashv} \vdash \Gamma$,
4. $\Sigma \mathrel{\vdash\!\!\!-\!\!\!-} \vdash \Gamma$.

The distinguished types of reasoning could be characterized dynamically.

1. (a) this reasoning starts with a set of asserted propositions which will give evidence;
   (b) a set of not asserted propositions is created, for the propositions will be given evidence;
2. (a) this reasoning starts with a set of asserted propositions for which will be given evidence;
   (b) a set of not asserted propositions is created, the propositions will give evidence;
3. (a) this reasoning starts with a set of propositions which are not asserted and will give evidence;
   (b) a set of asserted propositions is created, for the propositions will be given evidence;

[7] In another simple argument we may argue for assertion of the negation of this proposition. Thus the complex argument may be conceived as an argument for the suspension of this proposition. E.g., there are some arguments for the presence of general Błasik in the cockpit and there are some arguments against his presence in the cockpit. The arguments for are not convincing for me and the arguments against are not convincing for me. For these reasons I suspend the proposition that general Błasik was present in the cockpit. Another example: there are arguments for and there are arguments against the existence of a civilization outside Earth. Neither of the arguments is deciding. For this reason the proposition that there is a civilization outside Earth may be suspended.

4. (a) this reasoning starts with a set of not asserted propositions for which will be given evidence;
   (b) a set of asserted propositions is created, these propositions will give evidence.

The reasoning 1 is named *inference*. In the case of deductive reasoning the truth of conclusion is guaranteed by the truth of premisses. In the case of inductive reasoning it is conversely, namely the truth of premisses is guaranteed by the truth of conclusions (and enthymeme's premisses). In the case of analogy neither the truth of conclusions is guaranteed by the truth of premisses nor the truth of premisses is guaranteed by the truth of conclusions.

The reasoning 2 is called *explanation*. $\Gamma$ is the set of hypotheses. The hypothesis is used to explain facts stated by elements of $\Sigma$. This type of reasoning is an element of abduction. An abductive reasoning from $\Sigma$ to $\Gamma$ involves not simply a determination that, e.g., $\Gamma$ gives evidence for $\Sigma$, but also that $\Gamma$ is among the most economical explanations for $\Sigma$.

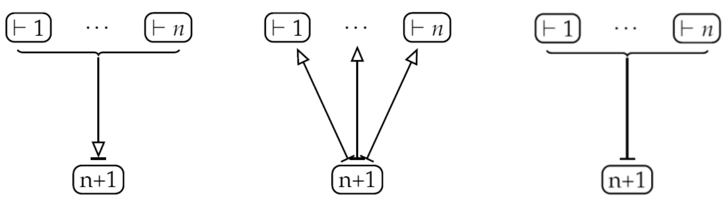
The reasoning 3 is named *verification*. The verification is used to confirm hypotheses. If the consequences of a hypothesis are confirmed, then the hypothesis is more probable.

The reasoning 4 is named *justification*. If from the set $\Gamma$ logically follows $\Sigma$, i.e. if the truth of $\Gamma$ guarantees the truth of $\Sigma$, then the justification is named *proof*. In mathematics a theorem $\phi$ is proved if and only if some already proved (asserted) theorems are found and it is shown that $\phi$ logically follows from these theorems.

Any premiss taken separately may give evidence (convergent argument) or to give evidence the premisses should be taken jointly (linked argument). The same is true about conclusions. The sign: $\underbrace{\qquad\qquad}$ or the sign: $\overbrace{\qquad\qquad}$ will be used to mark that propositions are taken jointly in the argument.

In diagrams to mark that $\Delta$ gives evidence for $\Lambda$ we will write: $\Delta \longrightarrow\!\!\!\mid \Lambda$. To mark that disjunction of propositions of $\Lambda$ logically follows from the conjunction of propositions of $\Delta$ we will write: $\Delta \longrightarrow \Lambda$. Instead of propositions in diagrams the numbers will be used.

There are three types of inference: deductive, inductive, analogical.




Deduction Induction Analogy

Using diagrams also other types of reasonings: explanation, verification and justification can be described.

The proposed description of arguments takes into account only pragmatic properties of propositions involved in argumentation and only logical relations between the set of premisses and the set of conclusions. It seems that the proposal is sufficiently rich to analyze a great variety of argumentations.

## References

Ajdukiewicz, K. (1965), Klasyfikacja rozumowań, *in* 'Język i poznanie', Vol. 2, PWN.

Berka, K. & Kreiser, L., eds (1971), Akademie-Verlag, Berlin. Zweite durchgesehene Auflage, 1973.

Bocheński, I. M. (1980), *Die zeitgenössischen Denkmethoden*, 8 edn, München.

Bocheński, J. M. (1992), *Współczesne metody myślenia*, Wydawnictwo «W drodze», Poznań.

Frege, G. (1879), *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle. Introduction dated: 18 XII 1878. Reprinted in (Frege 1993). Shortened version in (Berka & Kreiser 1971). English translation in (Heijenoort 1967, pp. 1–82).

Frege, G. (1993), *Begriffsschrift und andere Aufsätze*, Hildesheim.

Heijenoort, J. v., ed. (1967), *From Frege to Gödel. A Source Book in Mathematical Logic 1879–1931*, 1 edn, Harvard University Press, Cambridge. 2nd ed., 1971.

Reed, C., Walton, D. & Macagno, F. (2007), 'Argument diagramming in logic, law and artifical intelligence', *The Knowledge Engineering Review* **22**, 87–109. doi:10.1017/S0269888907001051.

Słupecki, J., Bryll, G. & Wybraniec-Skardowska, U. (1971), 'Theory of rejected propositions, I', *Studia Logica* **29**, 75–123.

Słupecki, J., Bryll, G. & Wybraniec-Skardowska, U. (1972), 'Theory of rejected propositions, I', *Studia Logica* **30**, 97–145.

Łukasiewicz, J. (1915), O nauce, *in* 'Poradnik dla samouków', 2 edn, Warszawa.

Kazimierz Trzęsicki
Chair of Logic, Informatics and Philosophy of Science
University of Białystok
kasimir@uwb.edu.pl

**Leila Amgoud and Florence Dupin de Saint-Cyr**
Institut de Recherche en Informatique de Toulouse (IRIT)

# ON THE QUALITY OF PERSUASION DIALOGS

**Abstract**: Several systems have been proposed for generating *persuasion dialogs* in which agents try to persuade each others to change their mind on a state of affairs. In this paper, we focus on the evaluation of the *quality* of those dialogs. We particularly propose three families of *measures*: i) measures of the quality of exchanged arguments, ii) measures of the behavior of each participating agent in terms of *coherence*, *aggressiveness* and the *novelty* of her arguments, iii) measures of the quality of the dialog itself in terms of *relevance* and *usefulness* of its moves. A notion of *conciseness* of a dialog is also introduced. For each persuasion dialog, we compute its *ideal* dialog which is a concise sub-dialog. The closer a dialog to its ideal sub-dialog, the better it is.

**Keywords**: argumentation, dialogue, measures of quality

## 1. Introduction

Persuasion is one of the main types of dialogs encountered in everyday life. It concerns two (or more) agents who disagree on a state of affairs, and each of them tries to persuade the others to change their minds. For that purpose, agents exchange arguments of different strengths. Several systems have been proposed in literature for allowing agents to engage in persuasion dialogs (e.g. [6, 7, 9, 10, 11, 12, 14]). A dialog system is built around three main components: i) a *communication language* specifying the locutions that will be used by agents during a dialog for exchanging information, arguments, etc., ii) a *protocol* specifying the set of rules governing the well-definition of dialogs such as who is allowed to say what and when? and iii) agents' strategies which are the different tactics used by agents for selecting their moves at each step in a dialog. It is worth mentioning that in these systems, only properties that are related to the protocol can be proved. Those properties are related to the way a dialog is generated. For instance, one can show whether a dialog terminates, or whether turn shifts equally between agents (if such rule is specified by the protocol), etc. However, a protocol does not say anything about the *quality* of the generated dialogs. Moreover, it is well-known that under the same protocol, different dialogs on the same subject may be generated. It is important to be able to compare them w.r.t. their quality. Such a comparison may help to refine the

ISBN 978–83–7431–305–6 ISSN 0860-150X 

protocols and to have more efficient ones. While there are numerous works on dialog protocols, no work is done on defining criteria for evaluating the persuasion dialogs generated under those protocols.

Besides, judging the properties of a dialog may be seen as a subjective issue. Two people listening to the same political debate may disagree on the "winner" and may have different feelings about the dialog itself.

In this paper, we investigate objective criteria for analyzing already generated dialogs whatever the protocol and the strategies that are used. We place ourselves in the role of an external observer who tries to evaluate a dialog, and we propose three families of measures: 1) Measures that evaluate the quality of exchanged arguments, 2) Measures that analyze the behavior of each participating agent in terms of *coherence* and *aggressiveness* in the dialog, and finally in terms of *borrowing* (when an agent uses arguments coming from other participating agents), 3) Measures of the properties of the dialog itself in terms of *relevance* and *usefulness* of its moves. A move is relevant if it does not deviate from the subject of the dialog, and it is useful if it is important to determine the outcome of the dialog. We propose also a criterion that evaluates the *conciseness* of a generated dialog. A dialog is concise if all its moves (i.e. the exchanged arguments) are both relevant to the subject and useful. Inspired by works on proof procedures that were proposed in the argumentation theory in order to check whether an argument is accepted or not [2], we compute and characterize a sub-dialog, called *ideal*, of the original one that is concise. The closer a dialog to its ideal sub-dialog, the better is its quality. All these measures are of great importance since they can be used as guidelines for generating the "best" dialogs. They can also serve as a basis for analyzing dialogs that held between agents.

The paper is organized as follows: Section 2 recalls the basics of the argumentation theory. Section 3 presents the basic concepts of a persuasion dialog. Section 4 describes the first family of measures, those evaluating arguments. Section 5 introduces measures that analyze the behavior of agents in a dialog. Section 6 presents the last family of measures, those devoted to the evaluation of a dialog. This paper unifies and develops the content of two previous works [3, 4].

## 2. Basics of argumentation systems

Argumentation is a reasoning model based on the construction and the comparison of arguments. Arguments are reasons for believing in statements, or for performing actions. In this paper, the origin of arguments is supposed to be unknown. In [8], an argumentation system is defined as follows:

**Definition 1** (Argumentation system)

An *argumentation system* is a pair $\mathsf{AS} = \langle \mathcal{A}, \mathcal{R} \rangle$, where $\mathcal{A}$ is a set of arguments and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is an attack relation. $(\alpha, \beta) \in \mathcal{R}$ means that argument $\alpha$ attacks $\beta$.

Note that to each argumentation system is associated a *directed graph* whose nodes are the different arguments, and the arcs represent the attack relation between them.

Since arguments are conflicting, it is important to know which arguments are acceptable. For that purpose different *acceptability semantics* have been proposed in [8]. In this paper, we only focus on *grounded* semantics. However, the work can be generalized to other semantics.

**Definition 2** (Defense-Grounded extension)

Let $\mathsf{AS} = \langle \mathcal{A}, \mathcal{R} \rangle$ and $\mathcal{E} \subseteq \mathcal{A}$.

- $\mathcal{E}$ *defends* an argument $\alpha \in \mathcal{A}$ iff $\forall\ \beta \in \mathcal{A}$, if $(\beta, \alpha) \in \mathcal{R}$, then $\exists \delta \in \mathcal{E}$ s.t. $(\delta, \beta) \in \mathcal{R}$.
- The *grounded extension* of $\mathsf{AS}$ is the least fixed point of a function $\mathcal{F}$ where $\mathcal{F}(\mathcal{E}) = \{\alpha \in \mathcal{A} | \mathcal{E} \text{ defends } \alpha\}$.

Each argumentation system has a unique grounded extension which may be empty. Moreover, when a system is finite (i.e. each argument is attacked by a finite number of arguments), its grounded extension is defined as follows: $\mathcal{E} = \bigcup_{i>0} \mathcal{F}^i(\emptyset)$. Depending on whether an argument belongs to this set or not, it is either accepted or rejected.

**Definition 3** (Argument status)

Let $\mathsf{AS} = \langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation system, and $\mathcal{E}$ its grounded extension. An argument $\alpha \in \mathcal{A}$ is *accepted* iff $\alpha \in \mathcal{E}$, it is *rejected* otherwise. We denote by $\texttt{Status}(\alpha, \mathsf{AS})$ the status of $\alpha$ in $\mathsf{AS}$.

**Proposition 1** ([2])

Let $\mathsf{AS} = \langle \mathcal{A}, \mathcal{R} \rangle$, $\mathcal{E}$ its grounded extension, and $\alpha \in \mathcal{A}$. If $\alpha \in \mathcal{E}$, then $\alpha$ is *indirectly defended*[1] by non-attacked arguments against all its attackers.

[1] An argument $\alpha$ is *indirectly defended* by $\beta$ iff there exists a finite sequence of distinct arguments $a_1, \ldots, a_{2n+1}$ such that $\alpha = a_1$, $\beta = a_{2n+1}$, and $\forall i \in [\![1, 2n]\!]$, $(a_{i+1}, a_i) \in \mathcal{R}$, $n \in \mathbb{N}^*$.

## 3. Persuasion dialogs

Throughout this section, $\mathcal{L}$ denotes a logical language. An *argument* is a reason for believing a statement. Thus, it has three main components: i) a *support* which is the set of premises on which the argument is grounded, it is thus a subset of $\mathcal{L}$, ii) a *conclusion* which is an element of $\mathcal{L}$ and iii) a *link* between the two.

**Notations**:

Support is a function which returns for each argument $\alpha$ its support, thus $\texttt{Support}(\alpha) \subseteq \mathcal{L}$. *arg* is a function which returns all the arguments that can be built from a subset $X$ of formulas ($X \subseteq \mathcal{L}$). Formulas is a function which returns the formulas included in the support of a set of arguments, hence if $A \subseteq arg(\mathcal{L})$, $\texttt{Formulas}(A) = \bigcup_{\alpha \in A} \texttt{Support}(\alpha)$.

Conflicts among arguments of $arg(\mathcal{L})$ are captured by a binary relation $\mathcal{R}_\mathcal{L}$ (i.e. $\mathcal{R}_\mathcal{L} \subseteq arg(\mathcal{L}) \times arg(\mathcal{L})$). We assume that each agent involved in a dialog recognizes any argument of $arg(\mathcal{L})$ and any conflict in $\mathcal{R}_\mathcal{L}$. This assumption does not mean that each agent is aware of all the arguments. But, it means that agents use the same logical language and the same definitions of argument and attack relation.

In what follows, a persuasion dialog consists of an exchange of arguments between two or more agents. The *subject* of such a dialog is an argument and its *aim* is to determine the status of that argument. Note that in [6], other kinds of moves (like questions, assertions) may be exchanged in a persuasion dialog. For our purpose, we consider only arguments since they allow us to determine the output of a dialog.

**Definition 4** (Move)

Let Ag be a set of symbols representing agents. A *move* $m$ is a triple $\langle S, H, \alpha \rangle$ such that:

- $S \in \texttt{Ag}$ is the agent that utters $m$, the function Speaker denotes this agent, i.e., $\texttt{Speaker}(m) = S$
- $H \subseteq \texttt{Ag}$ is the set of agents to which the move is addressed, the function Hearer denotes this set of agents: $\texttt{Hearer}(m) = H$
- $\alpha \in arg(\mathcal{L})$ is the content of the move, the function Content denotes the argument contained in the move: $\texttt{Content}(m) = \alpha$.

During a dialog several moves may be uttered. Those moves constitute a sequence denoted by $\langle m_1, \ldots, m_n \rangle$, where $m_1$ is the initial move whereas

$m_n$ is the final one. The empty sequence is denoted by $\langle\rangle$. These sequences are built under a given protocol like, for instance, the ones proposed in [6, 12]. For the purpose of our paper, we do not focus on particular protocols since we are not interested in generating dialogs but rather in analyzing a dialog which already took place.

**Definition 5** (Persuasion dialog)

A *persuasion dialog* $D$ is a non-empty and finite sequence of moves $\langle m_1, \ldots, m_n\rangle$ s.t. the *subject* of $D$ is $\texttt{Subject}(D) = \texttt{Content}(m_1)$, and the *length* of $D$, denoted $|D|$, is the number of moves: $n$. Each sub-sequence $\langle m_1, \ldots, m_i\rangle$ is a *sub-dialog* $D^i$ of $D$, denoted by $D^i \sqsubseteq D$.

An argumentation system is associated to each persuasion dialog in order to evaluate the status of its subject and that of each uttered argument.

**Definition 6** (AS of a persuasion dialog)

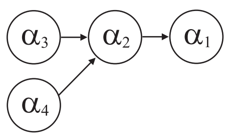
Let $D = \langle m_1, \ldots, m_n\rangle$ be a persuasion dialog. The *argumentation system* of $D$ is the pair $\mathsf{AS}_D = \langle \texttt{Args}(D), \texttt{Confs}(D)\rangle$ such that:

- $\texttt{Args}(D) = \{\texttt{Content}(m_i) | i \in [\![1, n]\!]\}$
- $\texttt{Confs}(D) = \{(\alpha, \beta) | \alpha, \beta \in \texttt{Args}(D) \text{ and } (\alpha, \beta) \in \mathcal{R}_{\mathcal{L}}\}$

To put it differently, $\texttt{Args}(D)$ and $\texttt{Confs}(D)$ return respectively the set of arguments exchanged in a dialog and the different conflicts among them.

**Example 1**

Let $D_1$ be a persuasion dialog between two agents $a_1$ and $a_2$ with $D_1 = \langle\langle a_1, \{a_2\}, \alpha_1\rangle, \langle a_2, \{a_1\}, \alpha_2\rangle, \langle a_1, \{a_2\}, \alpha_3\rangle, \langle a_1, \{a_2\}, \alpha_4\rangle, \langle a_2, \{a_1\}, \alpha_1\rangle\rangle$. The subject of $D_1$ is the argument $\alpha_1$. Let us assume the following conflicts among some of these arguments.




Thus, $\texttt{Args}(D_1) = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ and $\texttt{Confs}(D_1) = \{(\alpha_2, \alpha_1), (\alpha_3, \alpha_2), (\alpha_4, \alpha_2)\}$.

**Remark 1**

For any sub-dialog $D' \sqsubseteq D$, $\texttt{Args}(D') \subseteq \texttt{Args}(D)$ and $\texttt{Confs}(D') \subseteq \texttt{Confs}(D)$.

The *output* of a dialog is the status of the argument under discussion (i.e., the subject).

**Definition 7** (Output of a persuasion dialog)

Let $D$ be a persuasion dialog. The *output* of $D$, denoted by $\texttt{Output}(D)$, is $\texttt{Status}(\texttt{Subject}(D), \mathsf{AS}_D)$.

**Example 1** (Cont):

The grounded extension of $\mathsf{AS}_{D_1}$ is the set $\{\alpha_1, \alpha_3, \alpha_4\}$. Thus, $\alpha_1$ is accepted and consequently $\texttt{Output}(D_1) = \text{Accepted}$.

In the rest of the paper, we evaluate the quality of a given persuasion dialog $D$ according to three aspects:

1. the quality of the exchanged arguments
2. the behavior of each agent involved in the dialog
3. the conciseness of the dialog

We assume that the dialog $D$ is finite. Note that this assumption is not too strong since a main property of any protocol is the termination of the dialogs it generates [13]. A consequence of this assumption is that the argumentation system $\mathsf{AS}_D$ associated to $D$ is finite as well.

## 4. Measuring the quality of arguments

During a dialog, agents utter arguments that may have different *weights*. A weight may highlight the quality of information involved in the argument in terms, for instance, of certainty degree. It may also be related to the cost of revealing an information. In [1], several definitions of arguments' weights have been proposed, and their use for comparing arguments has been studied. It is worth noticing that the same argument may not have the same weight from one agent to another. In what follows, a weight in terms of a numerical value is associated to each argument. The greater this value is, the better the argument.

$$\texttt{weight} : arg(\mathcal{L}) \longrightarrow \mathbb{N}^*$$

The function $\texttt{weight}$ is given by *the agent who wants to analyze the dialog*. This agent may either be involved in the dialog or external. On the basis of arguments' weights, it is possible to compute the weight of a dialog as follows:

**Definition 8** (Measure of dialog weight)

Let $D$ be a persuasion dialog. The *weight* of $D$ is $\texttt{Weight}(D) = \sum_{\alpha \in \texttt{Args}(D)} \texttt{weight}(\alpha)$

**Property 1**

Let $D$ be a persuasion dialog. $\forall D' \sqsubseteq D$, $\texttt{Weight}(D') \leq \texttt{Weight}(D)$.

**Proof**

The result follows directly from Definition 8, the fact that $\texttt{Args}(D') \subseteq \texttt{Args}(D)$, and finally the fact that the function weight returns only positive values. ■

This measure allows to compare pairs of persuasion dialogs only on the basis of the exchanged arguments. It is even more interesting when the two dialogs have the same subject and got the same output.

It is also possible to compute the weight of arguments uttered by each agent in a given dialog. For that purpose, one needs to know what has been said by each agent. This can be computed by a simple projection on the dialog given that agent. Note that this projection is not usually a sub-dialog of $D$ (for instance, it may not contain $m_1$).

**Definition 9** (Dialog projection)

Let $D = \langle m_1, \ldots, m_n \rangle$ be a persuasion dialog and $a_i \in \texttt{Ag}$. The *projection* of $D$ on agent $a_i$ is $D^{a_i} = \langle m_{i_1}, \ldots, m_{i_k} \rangle$ such that $1 \leq i_1 \leq \ldots \leq i_k \leq n$ and $\forall l \in [1, k]$, $m_{i_l} \in D$ and $\texttt{Speaker}(m_{i_l}) = a_i$.

The contribution of each agent is defined as follows:

**Definition 10** (Measure of agent's contribution)

The *contribution* of an agent $a_i$ in a dialog $D$ is

$$\texttt{Contr}(a_i, D) = \frac{\sum_{\alpha_i \in \texttt{Args}(D^{a_i})} \texttt{weight}(\alpha_i)}{\texttt{Weight}(D)}$$

**Example 1** (Cont):

$D_1^{a_1} = \{\alpha_1, \alpha_3, \alpha_4\}$ and $D_1^{a_2} = \{\alpha_1, \alpha_2\}$. Suppose that an external agent who wants to analyze this dialog assigns the following weights to arguments: $\texttt{weight}(\alpha_1) = 1$, $\texttt{weight}(\alpha_2) = 4$, $\texttt{weight}(\alpha_3) = 2$ and $\texttt{weight}(\alpha_4) = 3$. Note that $\texttt{Weight}(D_1) = 10$. The contributions of the two agents are respectively $\texttt{Contr}(a_1, D_1) = 6/10$ and $\texttt{Contr}(a_2, D_1) = 5/10$.

Consider now an example in which an agent sends several times the same argument.

**Example 2**

Consider a persuasion dialog $D_2$ between two agents $a_1$ and $a_2$ with $\texttt{Args}(D_2) = \{\alpha, \beta\}$, $D_2^{a_1} = \{\alpha\}$ and $D_2^{a_2} = \{\beta\}$. Assume that there are 50 moves in $D_2$ of which 49 moves are uttered by agent $a_1$ and one move uttered by $a_2$. Assume also that an external agent assigns the following weights to arguments: $\texttt{weight}(\alpha) = 1$ and $\texttt{weight}(\beta) = 30$. The overall weight of the dialog is $\texttt{Weight}(D_2) = 31$. The contributions of the two agents are respectively $\texttt{Contr}(a_1, D_2) = 1/31$ and $\texttt{Contr}(a_2, D_2) = 30/31$.

It is easy to check that when the protocol under which a dialog is generated does not allow an agent to repeat an argument already given by another agent, then the sum of the contributions of the different agents is equal to 1.

**Property 2**

Let $D = \langle m_1, \ldots, m_n \rangle$ be a persuasion dialog and $a_1, \ldots, a_m$ the agents involved in $D$. $\sum_{i=1,\ldots,m} \texttt{Contr}(a_i, D) = 1$ iff $\nexists m_i, m_j$, $1 \leq i, j \leq n$, such that $\texttt{Speaker}(m_i) \neq \texttt{Speaker}(m_j)$ and $\texttt{Content}(m_i) = \texttt{Content}(m_j)$.

**Proof**

The proof follows directly from the definition. ■

As we will see in the next section, a more specific measure of contribution maybe defined if we focus on formulas that are involved in arguments. Indeed, contribution may be defined on the basis of formulas revealed by each agent. This requires to assign weights to formulas instead of arguments.

It is worth noticing that measure $\texttt{Contr}$ is *not monotonic* since the contribution of an agent may change during a dialog. However, at a given step of a dialog, the contribution of the agent who will present the next move will never decrease, whereas the contributions of the other agents may decrease.

**Proposition 2**

Let $D = \langle m_1, \ldots, m_n \rangle$ be a persuasion dialog, $a_i \in \texttt{Ag}$ and $m$ be a move such that $\texttt{Speaker}(m) = a_i$. It holds that $\texttt{Contr}(a_i, D \oplus m) \geq \texttt{Contr}(a_i, D)$ and $\forall a_j \in \texttt{Ag}$ with $a_j \neq a_i$, $\texttt{Contr}(a_j, D \oplus m) \leq \texttt{Contr}(a_j, D)$, with $D \oplus m = \langle m_0, \ldots, m_n, m \rangle$.

## 5. Analyzing the behavior of agents

The behavior of an agent in a given persuasion dialog may be analyzed on the basis of three main criteria: i) her degree of *aggressiveness* in the dialog, ii) the source of her arguments, i.e. whether she builds arguments using her own formulas, or rather the ones revealed by other agents, and finally iii) her degree of *coherence* in the dialog.

The first criterion, i.e. the aggressiveness of an agent in a dialog, amounts to computing to what extent an agent was attacking arguments sent by other agents. An aggressive agent prefers to destroy arguments presented by other parties rather than presenting arguments supporting her own point of view. Formally, the *aggressiveness degree* of an agent $a_i$ towards an agent $a_j$ during a persuasion dialog is equal to the number of its arguments that attack the other agent's arguments over the number of arguments it has uttered in that dialog.

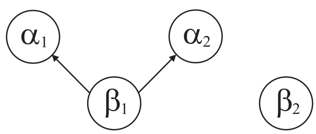
**Definition 11** (Measure of aggressiveness)

Let $D$ be a persuasion dialog and $a_i, a_j \in \mathtt{Ag}$. The *aggressiveness degree* of agent $a_i$ towards $a_j$ in $D$ is

$$\mathtt{Agr}(a_i, a_j, D) = \frac{|\{\alpha \in \mathtt{Args}(D^{a_i}) \text{ such that } \exists \beta \in \mathtt{Args}(D^{a_j}) \text{ and } (\alpha,\beta) \in \mathtt{Confs}(D)\}|}{|\mathtt{Args}(D^{a_i})|}^{2}.$$

**Example 3**

Let $D_3$ be a persuasion dialog between two agents $a_1$ and $a_2$. Assume that $\mathtt{Args}(D_3) = \{\alpha_1, \alpha_2, \beta_1, \beta_2\}$, $D_3^{a_1} = \{\alpha_1, \alpha_2\}$, $D_3^{a_2} = \{\beta_1, \beta_2\}$ and the conflicts are depicted in the figure below.




The aggressiveness degrees of the two agents are $\mathtt{Agr}(a_1, a_2, D_3) = 0$ and $\mathtt{Agr}(a_2, a_1, D_3) = 1/2$.

The aggressiveness degree of an agent changes as soon as a new argument is uttered by that agent. It decreases when that argument does not attack any argument of the other agent, and increases otherwise.

[2] The expression $|E|$ denotes the cardinal of the set E.

**Proposition 3**

Let $D = \langle m_1, \ldots, m_n \rangle$ be a persuasion dialog and $a_i, a_j \in \texttt{Ag}$. Let $m$ be a move such that $\texttt{Speaker}(m) = a_i$ and $D \oplus m = \langle m_1, \ldots, m_n, m \rangle$,

$$\texttt{Agr}(a_i, a_j, D \oplus m) \geq \texttt{Agr}(a_i, a_j, D) \text{ iff } \exists \alpha \in \texttt{Args}(D^{a_j}) \text{ such that } (\texttt{Content}(m), \alpha) \in \mathcal{R}_{\mathcal{L}}$$

The second criterion concerns the source of arguments. An agent can build her arguments either from her own knowledge base using her own formulas, or using formulas revealed by other agents in the dialog. In [5], this idea of borrowing formulas from other agents has been presented as one of the tactics used by agents for selecting the argument to utter at a given step of a dialog. The authors argue that by doing so, an agent minimizes the risk of being attacked subsequently. Let us now check to what extent an agent borrows information from other agents. Before that, let us first determine which formulas are owned by each agent according to what has been said in a dialog. Informally, a formula is owned by an agent if it is revealed for the first time by that agent. Note that a formula revealed for the first time by agent $a_i$ may also pertain to the base of another agent $a_j$ but, here, we are interested in *who reveals first* that formula.

**Definition 12** (Agent's formulas)

Let $D = \langle m_1, \ldots, m_n \rangle$ be a persuasion dialog and $a_i \in \texttt{Ag}$. The *formulas owned* by agent $a_i$ are: $\texttt{OwnF}(a_i, D) =$

$$\left\{ x \in \mathcal{L} \mid \exists m_j \text{ with } j \leq n \text{ and } \left| \begin{array}{l} \texttt{Speaker}(m_j) = a_i \text{ and } x \in \texttt{Support}(\texttt{Content}(m_j)) \\ \text{and } \nexists m_k \text{ with } k < j \text{ and } \left| \begin{array}{l} \texttt{Speaker}(m_k) \neq a_i \\ \text{and } x \in \texttt{Support}(\texttt{Content}(m_k)) \end{array} \right. \end{array} \right. \right\}$$

Now that we know which formulas are owned by each agent, we can compute the *degree of loan* of each agent. Note that from the strategical point of view, it is interesting to turn out an agent's argument against her in order to weaken her position. The borrowing degree can thus help for evaluating the strategical behavior of an agent.

**Definition 13** (Measure of loan)

Let $D$ be a persuasion dialog and $a_i, a_j \in \texttt{Ag}$. The *loan degree* of agent $a_i$ from agent $a_j$ in $D$ is:

$$\texttt{Loan}(a_i, a_j, D) = \frac{\left|\texttt{Formulas}(\texttt{Args}(D^{a_i})) \cap \texttt{OwnF}(a_j, D)\right|}{\left|\texttt{Formulas}(\texttt{Args}(D^{a_i}))\right|}.$$

It is worth mentioning that if agents do not borrow any formula from each other, then their contributions are independent. Hence, due to proposition 2, the sum of these contributions is equal to 1.

**Proposition 4**

Let $a_1, \ldots, a_m \in \texttt{Ag}$ be the agents involved in a persuasion dialog $D$. If $\forall i \neq j, \texttt{Loan}(a_i, a_j, D) = 0$, then $\sum_{i=1,\ldots,m} \texttt{Contr}(a_i, D) = 1$.

The third criterion concerns the coherence of an agent. Indeed, in a persuasion dialog where an agent $a_i$ defends her point of view, it is important to detect when this agent contradicts herself. There are two kinds of self contradiction:

1. an *explicit* contradiction in which an agent presents an argument and a counter-argument in the same dialog. Such conflicts appear in the argumentation system $\mathsf{AS}_{D^{a_i}} = \langle \texttt{Args}(D^{a_i}), \texttt{Confs}(D^{a_i}) \rangle$ associated with the moves uttered by agent $a_i$. Thus, the set $\texttt{Confs}(D^{a_i})$ is not empty.
2. an *implicit* contradiction appearing in a "complete" version of the agent's argumentation system.

The complete version of an argumentation system takes into account not only the set of arguments which are explicitly expressed in a dialog by an agent, i.e. $\texttt{Args}(D^{a_i})$, but also all the arguments that may be built from the set of formulas involved in the arguments of $\texttt{Args}(D^{a_i})$. Due to the monotonic construction of arguments, for any set $A$ of arguments, $A \subseteq arg(\texttt{Formulas}(A))$ but the reverse is not necessarily true. As a consequence, new conflicts may appear. This shows clearly that the argumentation system associated with a dialog is not necessarily "complete".

**Definition 14** (Complete AS)

The *complete AS* of a persuasion dialog $D$ is

$$\mathsf{CAS}_D = \langle arg(\texttt{Formulas}(\texttt{Args}(D))), \mathcal{R}_c \rangle$$

where $\mathcal{R}_c = \{(\alpha, \beta)$ such that $\alpha, \beta \in arg(\texttt{Formulas}(\texttt{Args}(D)))$ and $(\alpha, \beta) \in \mathcal{R}_{\mathcal{L}}\}$.

This definition is valid for any dialog projection $D^{a_i}$. Recall that $\texttt{Args}(D) \subseteq arg(\texttt{Formulas}(\texttt{Args}(D))) \subseteq arg(\mathcal{L})$ and $\texttt{Confs}(D) \subseteq \mathcal{R}_c \subseteq \mathcal{R}_{\mathcal{L}}$. Note also that the status of an argument $\alpha$ in a system $\mathsf{AS}_D$ is not necessarily the same in the complete system $\mathsf{CAS}_D$. The next definition evaluates to what extent an agent is incoherent in a dialog.

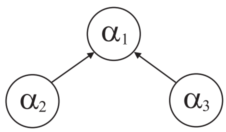
**Definition 15** (Measure of incoherence)

Let $D$ be a persuasion dialog, $a_i \in \mathtt{Ag}$ and $\mathsf{CAS}_{D^{a_i}} = \langle \mathcal{A}_c^{a_i}, \mathcal{R}_c^{a_i} \rangle$. The *incoherence degree* of agent $a_i$ in $D$ is

$$\mathtt{Inc}(a_i, D) = \frac{|\mathcal{R}_c^{a_i}|}{|\mathcal{A}_c^{a_i} \times \mathcal{A}_c^{a_i}|}.$$

**Example 4**

Let $D_4$ be a persuasion dialog in which agent $a_1$ has uttered two arguments $\alpha_1$ and $\alpha_2$. Let us assume that from the formulas of those arguments a third argument, say $\alpha_3$, is built. The figure below depicts the conflicts among the three arguments. The incoherence degree of agent $a_1$ is equal to 2/9.




Note that, the above definition is general enough to capture both explicit and implicit contradictions. Moreover, this measure is more precise than the one defined on the basis of attacked arguments, i.e. $\mathtt{Inc\_bis}(a_i, D) = \frac{|\{\beta \in \mathcal{A}_c^{a_i} \text{ such that } \exists (\alpha,\beta) \in \mathcal{R}_c^{a_i}\}|}{|\mathcal{A}_c^{a_i}|}$. Using this measure, the incoherence degree of agent $a_1$ is 1/3. Even if the argument $\alpha_1$ is attacked by two arguments, only one conflict is considered.

It is easy to check that if an agent is aggressive towards herself, then she is incoherent.

**Property 3**

Let $D$ be a persuasion dialog and $a_i \in \mathtt{Ag}$. If $\mathtt{Agr}(a_i, a_i, D) > 0$, then $\mathtt{Inc}(a_i, D) > 0$.

**Proof**

Let $D$ be a persuasion dialog and $a_i \in \mathtt{Ag}$. Assume that $\mathtt{Agr}(a_i, a_i, D) > 0$. This means that $\exists (\alpha, \beta) \in \mathtt{Confs}(D^{a_i})$. Consequently, $|\mathcal{R}_c^{a_i}| > 0$. This is due to the fact that $\mathtt{Confs}(D^{a_i}) \subseteq \mathcal{R}_c^{a_i}$. ■

The following example shows that the reverse is not always true.

**Example 5**

Let $D_5$ be a persuasion dialog and $a_i \in \mathtt{Ag}$. Assume that $\mathtt{Args}(D_5^{a_i}) = \{\alpha_1, \alpha_2\}$, and $\mathtt{Confs}(D_5^{a_i}) = \emptyset$. It means that $\mathtt{Agr}(a_i, a_i, D_5) = 0$. Suppose that $\mathsf{CAS}_{D_5^{a_i}} = \langle\ \{\alpha_1, \alpha_2, \alpha_3\}, \{(\alpha_3, \alpha_1), (\alpha_3, \alpha_2)\}\rangle$ is its associated complete argumentation system. It is clear that $\texttt{Inc}(a_i, D_5) = 2/9$.

Similarly, it can be shown that if agent $a_i$ is aggressive towards agent $a_j$ and if all the formulas of $a_i$ are borrowed from $a_j$, then $a_j$ is for sure incoherent. Note that $a_i$ might be coherent if she has not used conflicting arguments.

**Proposition 5**

Let $D$ be a persuasion dialog and $a_i, a_j \in \texttt{Ag}$. If $\texttt{Loan}(a_i, a_j, D) = 1$ and $\texttt{Agr}(a_i, a_j, D) > 0$, then $\texttt{Inc}(a_j, D) > 0$.

**Proof**

Let $\mathsf{CAS}_{D^{a_i}} = \langle \mathcal{A}_c^{a_i}, \mathcal{R}_c^{a_i}\rangle$ and $\mathsf{CAS}_{D^{a_j}} = \langle \mathcal{A}_c^{a_j}, \mathcal{R}_c^{a_j}\rangle$. It is clear that $\texttt{Loan}(a_i, a_j, D) = 1$ means that every formula used by $a_i$ has been first revealed by $a_j$, it implies that $\mathcal{A}_c^{a_i} \subseteq \mathcal{A}_c^{a_j}$ (1). Now if $\texttt{Agr}(a_i, a_j, D) > 0$ then it means that $\exists \alpha \in \texttt{Args}(D^{a_i})$ that is attacked by an argument of $\texttt{Args}(D^{a_j})$. From (1), we get that $\alpha \in \mathcal{A}_c^{a_j}$ hence $a_j$ is self-contradicting. ■

Note that incoherence is not necessarily a bad behavior, it depends on the aim of the participants: the goal may either be to win the debate whatever the other says or to discuss and take into account new information. In the last case, changing its opinion is a self-contradiction but may be a constructive attitude.

## 6. Measuring the conciseness of a dialog

It is very common that a dialog contains redundancies or useless moves. Thus, only some arguments may be useful for computing the output of the dialog. In this section, we are interested in characterizing the useful moves in a dialog and identifying the *ideal* version of a dialog. We start by presenting different criteria for evaluating each move in a dialog, then we provide a procedure for computing the ideal version of a given dialog.

### 6.1. Quality of moves

In everyday life, it is very common that agents deviate from the subject of the dialog. We first define a criterion that evaluates to what extent the moves uttered are in relation with the subject of the dialog. This amounts to check whether there exists a path from the argument presented by the agent towards the argument representing the subject in the graph of the argumentation system associated to the dialog.

**Definition 16** (Relevant and useful move)

Let $D = \langle m_1, \ldots, m_n \rangle$ be a persuasion dialog. A move $m_i$, with $i \in [\![1, n]\!]$, is *relevant* to $D$ iff there exists a path (not necessarily directed) from $\texttt{Content}(m_i)$ to $\texttt{Subject}(D)$ in the directed graph associated with $\mathsf{AS}_D$. A move $m_i$ is *useful* iff there exists a directed path from $\texttt{Content}(m_i)$ to $\texttt{Subject}(D)$ in this graph.

**Example 3** (Cont):

Assume that $\texttt{Subject}(D_3) = \alpha_1$. It is clear that $\alpha_3, \beta_1$ are relevant while $\beta_2$ is not and that $\beta_1$ is useful while $\alpha_3$ is not.

**Property 4**

If a move $m$ is useful in a dialog $D$, then $m$ is relevant to $D$.

**Proof**

If a move $m$ is useful then there exists a directed path from $\texttt{Content}(m)$ to $\texttt{Subject}(D)$, thus $m$ is relevant to $D$. ■

One can define a measure, called $\texttt{Relevance}(D)$, that computes the percentage of moves that are relevant in a dialog $D$[3]. In Example 3, $\texttt{Relevance}(D) = 3/4$. It is clear that the greater this degree is, the better the dialog. When the relevance degree of a dialog is equal to 1, this means that agents did not deviate from the subject of the dialog. Useful moves are those that have a more direct influence on the status of the subject. However, this does not mean that their presence has an impact on the output of the dialog. Moves that have a real impact on the status of the subject are called *decisive*.

**Definition 17** (Decisive move)

Let $D = \langle m_1, \ldots, m_n \rangle$ be a persuasion dialog and $\mathsf{AS}_D$ its argumentation system. A move $m_i$, with $i \in [\![1, n]\!]$, is *decisive* in $D$ iff

$$\texttt{Status}(\texttt{Subject}(D), \mathsf{AS}_D) \neq \texttt{Status}(\texttt{Subject}(D), \mathsf{AS}_D \ominus \texttt{Content}(m_i))$$

where $\mathsf{AS}_D \ominus \texttt{Content}(m_i) = \langle A', R' \rangle$ such that $A' = \texttt{Args}(D) \setminus \{\texttt{Content}(m_i)\}$ and $R' = \texttt{Confs}(D) \setminus \{(x, \texttt{Content}(m_i)), (\texttt{Content}(m_i), x) | x \in \texttt{Args}(D)\}$.

[3] $\texttt{Relevance}(D) = \frac{|\{m_{i=1,\ldots,n} \text{ such that } m_i \text{ is relevant to D}\}|}{|D|}$

It can be checked that if a move is decisive, then it is useful. This means that there exists a directed path from the content of this move to the subject of the dialog in the graph of the argumentation system associated to the dialog.

**Proposition 6**

If a move $m$ is decisive in a persuasion dialog $D$, then $m$ is useful in $D$.

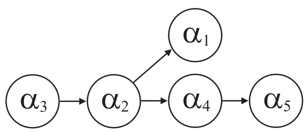
**Proof**

Assume that $m$ is a decisive move in $D$ and that $\texttt{Subject}(D)$ is accepted in $\mathsf{AS}_D$. According to Proposition 1, for any attacker of $\texttt{Subject}(D)$, $\texttt{Subject}(D)$ is indirectly defended by a non-attacked argument. Since $m$ is decisive, $\texttt{Subject}(D)$ is rejected in $\mathsf{AS}_D \ominus \texttt{Content}(m)$. This means that at least one attacker is no more indirectly defended by a non-attacked argument. Hence, removing $\texttt{Content}(m)$ eliminates a path from a non-attacked argument to this attacker. Hence $\texttt{Content}(m)$ is useful. If $\texttt{Subject}(D)$ is rejected in $\mathsf{AS}_D$ and accepted in $\mathsf{AS}_D \ominus \texttt{Content}(m)$. This means that every attacker is defended by a non-attacked argument in $\mathsf{AS}_D \ominus \texttt{Content}(m)$. Hence the deletion of $\texttt{Content}(m)$ has eliminated every direct or indirect attacker of the subject. This means that $\texttt{Content}(m)$ was on a path from an attacker to the subject hence it was useful in $D$. ■

From Property 4, it follows that each decisive move is also relevant. Note that the converse is not true as shown in the following example.

**Example 6**

Let $D_6$ be a dialog whose subject is $\alpha_1$ and whose graph is the following:




The grounded extension of $\mathsf{AS}_{D_6}$ is $\{\alpha_1, \alpha_3, \alpha_5\}$. It is clear that the argument $\alpha_4$ is relevant to $\alpha_1$, but it is not decisive for $D_6$. Indeed, the removal of $\alpha_4$ will not change the status of $\alpha_1$ which is accepted.

The converse of Proposition 6 is not true since useful moves may not be decisive:

**Example 7**

Let $D_7$ be a dialog whose argumentation system is the one given in Example 4 and whose subject is $\alpha_1$. Note that neither $\alpha_2$ nor $\alpha_3$ is decisive in $D_7$. However, this does not mean that the two arguments should be removed since the status of $\alpha_1$ depends on at least one of them (they are both useful).

On the basis of the above notion of decisiveness of moves, we can define *the degree of decisiveness* of the entire dialog as the percentage of moves that are decisive.

### 6.2. Canonical dialogs

As shown in the previous sub-section, some moves may not be important in a dialog and removing them does not have any impact on the output of the dialog. In this section, we characterize sub-dialogs, called *canonical*, which return the same output as an original dialog. In [2], a proof procedure that tests the membership of an argument to a grounded extension has been proposed. The basic notions of this procedure are revisited and adapted for the purpose of characterizing canonical dialogs.

**Definition 18** (Dialog branch)

Let $D$ be a persuasion dialog and $\mathsf{AS}_D = \langle \texttt{Args}(D), \texttt{Confs}(D) \rangle$ its argumentation system. A *dialog branch* for $D$ is a sequence $\langle \alpha_0, \ldots, \alpha_p \rangle$ of arguments such that $\forall i, j \in [\![0, p]\!]$

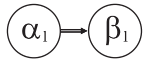
1. $\alpha_i \in \texttt{Args}(D)$
2. $\alpha_0 = \texttt{Subject}(D)$
3. if $i \neq 0$ then $(\alpha_i, \alpha_{i-1}) \in \texttt{Confs}(D)$
4. if $i$ and $j$ are even and $i \neq j$ then $\alpha_i \neq \alpha_j$
5. if $i$ is even and $i \neq 0$ then $(\alpha_{i-1}, \alpha_i) \notin \texttt{Confs}(D)$
6. $\forall \beta \in \texttt{Args}(D)$, $\langle \alpha_0, \ldots, \alpha_p, \beta \rangle$ is not a dialog branch for $D$.

Intuitively, a dialog branch is a kind of partial sub-graph of $\mathsf{AS}_D$ in which the nodes contains arguments and the arcs represent inverted conflicts. Note that arguments that appear at even levels are not allowed to be repeated. Moreover, these arguments should strictly attack[4] the preceeding argument. The last point requires that a branch is maximal. Let us illustrate this notion with examples.

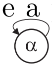
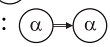
[4] An argument $\alpha$ strictly attacks an argument $\beta$ in a argumentation system $\langle \mathcal{A}, \mathcal{R} \rangle$ iff $(\alpha, \beta) \in \mathcal{R}$ and $(\beta, \alpha) \notin \mathcal{R}$.

**Example 3** (Cont):

The only dialog branch that can be built from dialog $D_3$ is:




**Example 8**

Let $D_8$ be a persuasion dialog whose subject is $\alpha$ and whose graph is the following:  The only possible dialog branch associated to this dialog is the following: 

**Proposition 7**

A dialog branch is non-empty and finite.

**Proof**

– A dialog branch is non-empty since the subject of the original persuasion dialog belongs to the branch.

– Let us assume that there exists an infinite dialog branch for a given persuasion dialog $D$. This means that there is an infinite sequence $\langle \alpha_0, \alpha_1, \ldots \rangle$ that forms a dialog branch. In this sequence, the number of arguments of even index and of odd index are infinite. According to Definition 5, the persuasion dialog $D$ is finite, thus both sets $\texttt{Args}(D)$ and $\texttt{Confs}(D)$ are finite. Consequently, the set of arguments that belong to the sequence $\langle \alpha_0, \alpha_1, \ldots \rangle$ is finite. Hence, there is at least one argument that is repeated at an even index. This is impossible. ■

Moreover, it is easy to check the following result:

**Proposition 8**

For each dialog branch $\langle \alpha_0, ..., \alpha_k \rangle$ of a persuasion dialog $D$ there exists a unique directed path $(\alpha_k, \alpha_{k-1}, ..., \alpha_0)$ of same length[5] $(k)$ in the directed graph associated to $\mathsf{AS}_D$.

**Proof**

Let $\langle \alpha_0, ..., \alpha_k \rangle$ be a dialog branch for $D$, from Definition 18.3, it follows that $\forall i \in [\![1, k]\!]$, $(\alpha_i, \alpha_{i-1}) \in Confs(D)$. Hence there is a path of length $k$ in $\mathsf{AS}_D$ from $\alpha_k$ to $\alpha_0$. From Definition 18.2, $\alpha_0 = Subject(D)$. ■

In what follows, we show that when a dialog branch is of even-length, then its leaf is not attacked in the original dialog.

[5] The length of a path is defined by its number of arcs.

**Theorem 1**

$\langle \alpha_0, .., \alpha_p \rangle$ being a dialog branch for $D$, if $p$ is even then $\nexists \beta \in \texttt{Args}(D)$ such that $(\beta, \alpha_p) \in \texttt{Confs}(D)$

**Proof**

If $\exists \beta \in \texttt{Args}(D)$ such that $(\beta, \alpha_p) \in \texttt{Confs}(D)$ then a new sequence beginning by $\langle \alpha_0, \ldots \alpha_p, \beta \rangle$ would be a dialog branch, which is forbidden by Definition 18.6. ■

Let us now introduce the notion of a dialog tree.

**Definition 19** (Dialog tree)

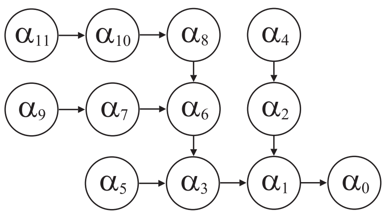
A *dialog tree* of $D$, denoted by $D^t$, is a finite tree whose branches are all the possible dialog branches that can be built from $D$.

We denote by $\mathsf{AS}_{D^t}$ the argumentation system associated to $D^t$, $\mathsf{AS}_{D^t} = \langle A^t, C^t \rangle$ such that $A^t = \{\alpha \in \texttt{Args}(D)$ such that $\alpha$ appears in a node of $D^t\}$ and $C^t = \{(\alpha, \beta) \in \texttt{Confs}(D)$ such that $(\beta, \alpha)$ is an arc of $D^t\}$.

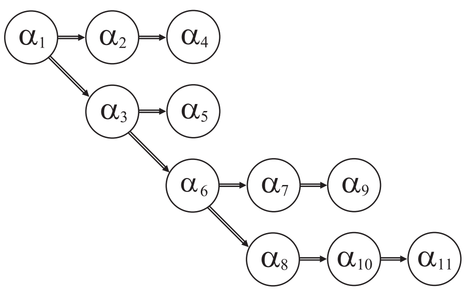
Hence, a dialog tree is a tree whose root is the subject of the persuasion dialog.

**Example 9**

Let us consider $D_9$ whose subject is $\alpha_1$ and whose graph is the following:



The dialog tree associated to this dialog is:



Note that the argument $\alpha_0$ does not belong to the dialog tree.

**Proposition 9**

Each persuasion dialog has exactly one corresponding dialog tree.

**Proof**

This follows directly from the definition of the dialog tree. Indeed, the root of the tree is the subject of the persuasion dialog. Moreover, all the possible branches are considered. ■

An important result states that the status of the subject of the original persuasion dialog $D$ is exactly the same in both argumentation systems $\mathsf{AS}_D$ and $\mathsf{AS}_{D^t}$ (where $\mathsf{AS}_{D^t}$ is the argumentation system whose arguments are all the arguments that appear in the dialog tree $D^t$ and whose attacks are obtained by inverting the arcs between those arguments in $D^t$).

**Theorem 2**

$\mathtt{Status}(\mathtt{Subject}(D), \mathsf{AS}_D) = \mathtt{Status}(\mathtt{Subject}(D), \mathsf{AS}_{D^t})$.

**Proof**

The proof of this theorem is based on two theorems given farther that are referring to the notion of canonical tree.

- If $\mathtt{Subject}(D)$ is accepted in $\mathsf{AS}_D$. then using Theorem 4 we get that there exists a canonical tree $D_i^c$ such that $\mathtt{Subject}(D)$ is accepted in $\mathsf{AS}_{D_i^c}$. Moreover, the way $D_i^c$ has been constructed (by an AND/OR process) imposes that $D_i^c$ contains every direct child of the subject in $D^t$. Furthermore, Theorem 3 shows that every branch of $D_i^c$ is of even length. Every leaf of this canonic tree, by definition, is non-attacked in $D_i^c$ and by definition in $\mathsf{AS}_{D^t}$. Using Definition 18.4 we get that in each branch of $\mathsf{AS}_{D^t}$, each even node strictly attacks the previous node. Hence, by construction, for each direct attacker of the subject in $\mathsf{AS}_{D^t}$, there exists at least one defender non-attacked in $\mathsf{AS}_{D^t}$ (leaf of $D_i^c$), the defense being strict, the subject belongs to the basic extension of $\mathsf{AS}_{D^t}$.
- If $\mathtt{Subject}(D)$ is accepted in $\mathsf{AS}_{D^t}$ then there exists a non-attacked defender against every direct attacker of the subject in $\mathsf{AS}_{D^t}$. This means that there exists a canonical tree based on $\mathsf{AS}_{D^t}$ having only even length branches. The subject is accepted in this canonical tree using Theorem 3, which implies that the subject is accepted in $D$ using Theorem 4. ■

In order to compute the status of the subject of a dialog, we can consider the dialog tree as an And/Or tree. A node of an even level is an And node, whereas a node of odd level is an Or one. This distinction between nodes is due to the fact that an argument is accepted if it can be defended against all its attackers. A dialog tree can be decomposed into one or several trees called canonical trees. A *canonical tree* is a subtree of $D^t$ whose root is $\texttt{Subject}(D)$ and which contains all the arcs starting from an even node and exactly one arc starting from an odd node.

**Definition 20** (Canonical tree)

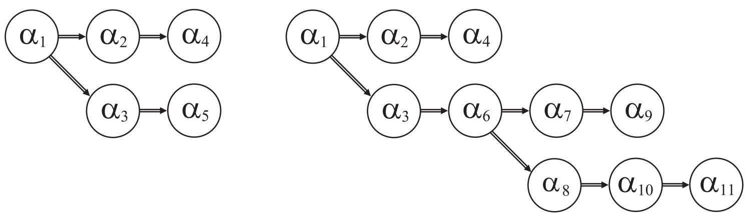
Let $D$ be a persuasion dialog, and let $D^t$ its dialog tree. $D^c$ is a *canonical tree* of $D^t$ if it is a subtree of $D^t$ built by levels as follows:

- $\texttt{Subject}(D)$ is its root (of level 0)
- and inductively:
  - if $\alpha$ is a node of even level in $D^c$ then for **every** $\beta \in D^t$ such that $(\alpha, \beta) \in D^t$, the node $\beta$ and the arc $(\alpha, \beta)$ is added to $D^c$.
  - if $\alpha$ is a node of odd level in $D^c$ and if $\alpha$ has at least one attacker in $D^t$ then for **exactly one** $\beta \in D^t$ such that $(\alpha, \beta) \in D^t$, the node $\beta$ and the arc $(\alpha, \beta)$ is added to $D^c$.

It is worth noticing that from a dialog tree one may extract at least one canonical tree. Let $D_1^c, \ldots, D_m^c$ denote those canonical trees. We will denote by $\mathsf{AS}_1^c, \ldots, \mathsf{AS}_m^c$ their corresponding argumentation systems. It can be checked that the status of $\texttt{Subject}(D)$ is not necessarily the same in these different systems.

**Example 10**

From the dialog tree of $D_9$, two canonical trees can be extracted:




It can be checked that the argument $\alpha_1$ is accepted in the argumentation system of the canonical tree on the left while it is rejected in the one of the right.

The following result characterizes the status of $\texttt{Subject}(D)$ in the argumentation system $\mathsf{AS}_i^c$ associated to a canonical tree $D_i^c$.

**Theorem 3**

Let $D$ be a persuasion dialog, $D_i^c$ a canonical tree and $\mathsf{AS}_i^c$ its corresponding argumentation system. $\mathtt{Subject}(D)$ is accepted in $\mathsf{AS}_i^c$ iff all the branches of $D_i^c$ are of even-length.

**Proof**

Let $D$ be a persuasion dialog, $D_i^c$ a canonical tree and $\mathsf{AS}_i^c$ its corresponding argumentation system.

- Assume that $\mathtt{Subject}(D)$ is accepted in $\mathsf{AS}_i^c$, and that there is a branch of $D_i^c$ whose length is odd. This means that the leaf of this branch, say $\alpha$, indirectly attacks $\mathtt{Subject}(D)$ (the root of the branch).
  - Either $\alpha$ is not attacked in $\mathsf{AS}_i^c$ it means that $\alpha$ is accepted hence the second node of the branch is a direct attacker of $\mathtt{Subject}(D)$ that is not defended by a non attacked argument, i.e., $\mathtt{Subject}(D)$ would not be accepted in $\mathsf{AS}_i^c$.
  - Either $\alpha$ is attacked in $\mathsf{AS}_i^c$ then it can only be attacked by an argument already present in the branch (hence itself attacked), else the branch would not satisfied Definition 18.6. This also means that the second node of the branch is a direct attacker of $\mathtt{Subject}(D)$ that is not defended by a non attacked argument.
- Assume now that all the branches of $D_i^c$ are of even length, then for each branch the leaf is accepted since it is not attacked in $\mathsf{AS}_i^c$ (using Theorem 1). Then iteratively considering each even node from the leaf to the root, they can all be added to the grounded extension since the leaf defends the penultimate even node against the attack of the last odd node and so on and by construction for each odd node attacking an even node there is a deeper even node that strictly defends it (due to Definition 18.5). Hence each even node is in the grounded extension, so $\mathtt{Subject}(D)$ is accepted in $\mathsf{AS}_i^c$ ∎

The following result follows immediately from this Theorem and Theorem 1.

**Corollary 1**

Let $D$ be a persuasion dialog, $D_i^c$ a canonical tree and $\mathsf{AS}_i^c$ its corresponding argumentation system. If $\mathtt{Subject}(D)$ is accepted in $\mathsf{AS}_i^c$, then all the leaves of $D_i^c$ are not attacked in $D$.

**Proof**

According to Theorem 3, since $\mathtt{Subject}(D)$ is accepted in $\mathsf{AS}_i^c$, then all its branches are of even-length. According to Theorem 1, the leaf of each branch of even-length is an argument that is not attacked in $D$. Thus, all the leaves of $D^c_i$ are not attacked in $D$. ■

An important result shows the link between the outcome of a dialog $D$ and the outcomes of the different canonical trees.

**Theorem 4**

Let $D$ be a persuasion dialog, $D^c_1$, …, $D^c_m$ its different canonical trees and $\mathsf{AS}^c_1, \ldots, \mathsf{AS}^c_m$ their corresponding argumentation systems. $\texttt{Output}(D)$[6] is accepted iff $\exists\, i \in [\![1, m]\!]$ such that $\texttt{Status}(\texttt{Subject}(D), \mathsf{AS}^c_i)$ is accepted.

**Proof**

Let $D$ be a persuasion dialog, $D^c_1$, …, $D^c_m$ its different canonical trees and $\mathsf{AS}^c_1, \ldots, \mathsf{AS}^c_m$ their corresponding argumentation systems.

- Let us assume that there exists $D^c_j$ with $1 \leq j \leq m$ and $\texttt{Status}(\texttt{Subject}(D), \mathsf{AS}^c_j)$ is accepted. According to Theorem 3, this means that all the branches of $D^c_j$ are of even length. From Corollary 1, it follows that the leaves of $D^c_j$ are all not attacked in the graph of the original dialog $D$.
  Let $2i$ be the depth of $D^c_j$ (i.e. the maximum number of moves of all dialog branches of $D^c_j$).
  We define the height of a node $N$ in a tree as the depth of the sub-tree of root $N$.
  We show by induction on $p$ that $\forall$ p such that $0 \leq p \leq i$, the set $\{y | y$ is an argument of even indice and in a node of height $\leq 2p$ belonging to $D^c_j\}$ is included in the grounded extension of $\mathsf{AS}_D$).
  - Case $p = 0$. The leaves of $D^c_j$ are not attacked in D (according to Corollary 1). Thus, they belong to the grounded extension of $\mathsf{AS}_D$.
  - Assume that the property is true to an order $p$ and show that it is also true to the order $p+1$. It is sufficient to consider the arguments that appear at even levels and in a node of height $2p + 2$ of $D^c_j$. Let $y$ be such an argument. Since $y$ appears at an even level, then all the arguments $y'$ attacking $y$ in $\mathsf{AS}_D$ appear in $D^c_j$ as children of $y$ (otherwise the branch would not be maximal or $D^c_j$ would not be canonic), and each $y'$ is itself strictly attacked in $\mathsf{AS}_D$ by exactly one argument $z$ appearing in $D^c_j$ as a child of $y'$. Thus, each $z$ is at an even level in $D^c_j$ and appears as a node of height $2p$ of $D^c_j$. By induction hypothesis, each argument $z$ is in the grounded

[6] Recall that $\texttt{Output}(D) = \texttt{Status}(\texttt{Subject}(D), \mathsf{AS}_D)$.

extension of $\mathsf{AS}_D$. Since all attackers of $y$ have been considered, thus the grounded extension of $\mathsf{AS}_D$ defends $y$. Consequently $y$ is also in this grounded extension.

- Let us assume that $\texttt{Status}(\texttt{Subject}(D), \mathsf{AS}_D)$ is accepted. Let $i_0$ be the smallest index $\geq 0$ such that $\texttt{Subject}(D) \in \mathcal{F}^{i_0}(\mathcal{C}^7)$. Let us show by induction on $i$ that if an argument $\alpha \in \texttt{Args}(D)$ is in $\mathcal{F}^i(\mathcal{C})$ then there exists a canonical tree of root $\alpha$ for $D$[8] having a depth $\leq 2i$ and having only branches of even length.
  - Case $i = 0$: if $\alpha \in \mathcal{C}$, then $\alpha$ itself is a canonical tree of root $\alpha$ and depth 0.
  - Assume that the property is true at order $i$ and consider the order $i+1$. Hence, let us consider $\alpha \in \mathcal{F}^{i+1}(\mathcal{C})$ and $\alpha \notin \mathcal{F}^k(\mathcal{C})$ with $k < i+1$.

    Let $x_1, \ldots, x_n$ be the attackers of $\alpha$. Consider an attacker $x_j$. $x_j$ attacks $\alpha$, and $\alpha \in \mathcal{F}^{i+1}(\mathcal{C}) = \mathcal{F}(\mathcal{F}^i(\mathcal{C}))$. According to Proposition 4.1 in [2], there exists $y$ in the grounded extension of $\mathsf{AS}_D$ such that $y$ attacks strictly $x_j$. Since $y$ defends $\alpha$ (definition of $\mathcal{F}$) then $y \in \mathcal{F}^i(\mathcal{C})$. By induction hypothesis applied to $y$, there exists a canonical tree whose root is $y$ and the depth is $\leq 2i$. The same construction is done for each $x_j$. So we get a canonical tree whose root is $\alpha$ and its depth is $\leq 2(i+1)$ and in which each branch has still an even length.

  Now, from the fact that $\texttt{Subject}(D) \in \mathcal{F}^{i_0}(\mathcal{C})$ we conclude that it exists a canonical tree of root $\texttt{Subject}(D)$ having each branch of even length. Using Theorem 3, we get that $\texttt{Subject}(D)$ is accepted in this canonical tree. ■

This result is of great importance since it shows that a canonical tree whose branches are all of even-length is sufficient to reach the same outcome as the original dialog in case the subject is accepted. When the subject is rejected, the whole dialog tree is necessary to ensure the outcome.

**Example 9** (Cont):

The subject $\alpha_1$ of dialog $D_9$ is accepted since there is a canonical tree whose branches are of even length (it is the canonical tree on the left in

---

[7] The set $\mathcal{C}$ contains all the arguments that are not attacked in $D$.

[8] Here, we consider a "canonical tree of root $\alpha$ for a dialog $D$". Its definition is more general than canonical tree for a dialog $D$ since it does not requires that all the branches start from the subject of the dialog (modifying item 2 of Definition 18) but requires that all the branches start from the node $\alpha$.

Example 10). It can also be checked that $\alpha_1$ is in the grounded extension $\{\alpha_1, \alpha_4, \alpha_5, \alpha_8, \alpha_9, \alpha_{11}\}$ of $\mathsf{AS}_{D_9}$.

So far, we have shown how to extract from a graph associated with a dialog its canonical trees. These canonical trees contain only useful (hence relevant) moves:

**Theorem 5**

Let $D_i^c$ be a canonical tree of a persuasion dialog $D$. Any move built on an argument of $D_i^c$ is *useful* in the dialog $D$.

**Proof**

By construction of $D_i^c$, there is a path in this tree from the root to each argument $\alpha$ of the canonical tree. According to Proposition 8, we get that there exists a corresponding directed path in $\mathsf{AS}_D$ from $\alpha$ to $\texttt{Subject}(D)$, hence a move containing the argument $\alpha$ is useful in $D$. ■

The previous theorem gives an upper bound of the set of moves that can be used to build a canonical tree, a lower bound is the set of decisive moves.

**Theorem 6**

Every argument of a decisive move belongs to the dialog tree and to each canonical tree.

**Proof**

If a move $m$ is decisive then, as seen in the proof of proposition 6,

- if the subject is accepted in $\mathsf{AS}_D$ then it exists at least a direct attacker of the subject that is no more inderectly defended by a non attacked argument in $\mathsf{AS}_D \ominus \texttt{Content}(m)$. The subject being accepted in $AD_D$, this means that there is a canonical tree having only branches of even length (according to Theorem 3). By construction, this canonic tree contains every direct attacker of the subject. If $\texttt{Content}(m)$ does not belong to this canonic tree then there is a defender of the subject on a path that does not contain $\texttt{Content}(m)$ in $\mathsf{AS}_D$, if it is the case for every direct attacker of the subject then the subject should have been accepted in $\mathsf{AS}_D \ominus \texttt{Content}(m)$. This is not possible, hence $\texttt{Content}(m)$ belongs to the canonical tree that accepts the subject.
- if the subject is rejected in $\mathsf{AS}_D$ but accepted in $\mathsf{AS}_D \ominus \texttt{Content}(m)$ then there exists a canonical tree hwhere all the branches are of even length in $\mathsf{AS}_D \ominus \texttt{Content}(m)$. Since the adding of /*content*$(m)$ leads to reject the subject, it means that $\texttt{Content}(m)$ attacks at least one direct

or indirect defender of the subject belonging to each canonical tree that accepts the subject in $\mathsf{AS}_D \ominus \texttt{Content}(m)$. The sequence containing the branch from the subject to that defender can be prolongated with $\texttt{Content}(m)$ in order to form a new branch of odd length in $\mathcal{D}^t$. Hence for every canonical tree that rejects the subject, $\texttt{Content}(m)$ has to belong one of their branch. ■

The converse is false since many arguments are not decisive. It is illustrated in Example 7, there are two attackers that are not decisive but the dialog tree contains both of them (as does the only canonical dialog for this example).

### 6.3. The ideal dialog

In the previous section, we have shown that from each dialog, a dialog tree can be built. This dialog tree contains direct and indirect attackers and defenders of the subject. From this dialog tree, interesting subtrees can be extracted and are called canonical trees. A canonical tree is a subtree containing only particular entire branches of the dialog tree (only one argument in favor of the subject is chosen for attacking an attacker while each argument against a defender is selected). In case the subject of the dialog is accepted it has been proved that there exists at least one canonical tree such that the subject is accepted in its argumentation system. This canonical tree is a candidate for being an ideal tree since it is sufficient to justify the acceptance of the subject against any attack available in the initial dialog. Among all these candidates, we define the ideal tree as the smallest one. In the case the subject is rejected in the initial dialog, then the dialog tree contains all the reasons to reject it, hence we propose to consider the dialog tree itself as the only ideal tree.

**Definition 21** (ideal trees and dialogs)

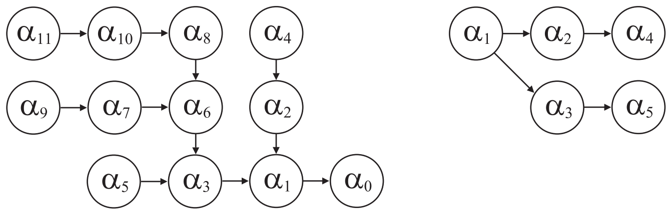
If a dialog $D$ has an accepted output

- then an *ideal tree* associated to $D$ is a canonical tree of $D$ in which $\texttt{Subject}(D)$ is accepted and having a minimal number of nodes among all the canonical graphs that also accept $\texttt{Subject}(D)$
- else the *ideal tree* is the dialog tree of $D$.

A dialog using once each argument of an ideal graph is called an *ideal dialog.*

**Example 9** (Cont):

An ideal Dialog for Dialog $D_9$ (on the left) has the following graph (on the right):




Given the above definition, an ideal dialog contains exactly the same number of moves that the number of nodes of the ideal graph.

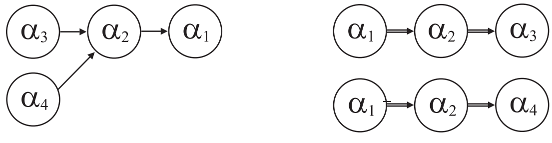
**Proposition 10**

Given a dialog $D$ whose subject is accepted. An ideal dialog $ID$ for $D$ is the shortest dialog with the same output, and such that every argument in favor of the subject in $ID$ (including $Subject(D)$ itself) is defended against any attack (existing in $D$).

**Proof**

If the subject is accepted in $D$ then, by construction, a canonical graph of $D$ contains every argument existing in $D$ that directly attacks the subject since they belong to all the possible dialog branches that can be built from $D$. But for any of them it contains only one attacker that is in favor of the subject (this attacker is a son of an "OR" node in the dialog tree), for each chosen argument in favor of the subject, all the attackers are present in the canonical tree (they are the sons of an "AND" node in the dialog tree). Moreover, if the subject is accepted then every branch of the canonical graph is of even length. It means that the leafs are in favor of the subject and not attacked in the initial dialog $D$. This property is true for any canonical graph. Then since the ideal dialog corresponds to the smallest canonical graph it means that it is the shortest dialog that satisfies this property. ∎

This property ensures that, when the subject is accepted in the initial dialog $D$, an ideal dialog $ID$ is the more concise dialog that entails an acceptation. In other words, we require that the ideal dialog should contain a set of arguments that sumarize $D$. Note that the ideal dialog exists but is not always unique. Here is an example of an argumentation system of a dialog which leads to two ideal trees (hence it will lead to at least two ideal dialogs).



So far, we have formally defined the notion of ideal dialog, and have shown how it is extracted from a persuasion dialog. It is clear that the closer (in terms of set-inclusion of the exchanged arguments) the dialog from its ideal version, the better the dialog.

## 7. Conclusion

Several systems have been proposed in literature for allowing agents to engage in persuasion dialogs. Different dialog protocols have then been discussed. These latter are the high level rules that govern a dialog. Examples of such rules are 'how the turn shifts between agents', and 'how moves are chained in a dialog'. All these rules should ensure 'correct' dialogs, i.e. dialogs that terminate and reach their goals. However, they do not say anything on the quality of the dialogs. One even wonders whether there are criteria for measuring the quality of a dialog. In this paper, we argue that the answer to this question is yes. Indeed, under the same protocol, different dialogs on the same subject may be generated, and some of them may be judged better than others. There are three kinds of reasons, each of them is translated into quality measures: i) the exchanged arguments are stronger, ii) the behavior of agents was 'ideal'. iii) the generated dialogs are more concise (i.e. all the uttered arguments have an impact on the result of the dialog). In this paper, the behavior of an agent is analyzed on the basis of three main criteria: its degree of aggressiveness, its degree of loan, and its degree of coherence.

We have also proposed three criteria for evaluating the moves of a persuasion dialog with respect to its subject: relevance, usefulness and decisiveness. Relevance only expresses that the argument of the move has a link with the subject (this link is based on the attack relation of the argumentation system). Usefulness is a more stronger relevance since it requires a directed link from the argument of the move to the subject. Decisive moves have a heavier impact on the dialog, since their omission changes the output of the dialog.

Inspired by works on proof theories for grounded semantics in argumentation, we have defined a notion of "ideal dialog". More precisely, we have first defined a dialog tree associated to a given dialog as the graph that contains every possible direct and indirect attackers and defenders of the subject. From this dialog tree, it is then possible to extract sub-trees called "ideal trees" that are sufficient to prove that the subject is accepted or rejected in the original dialog and this, against any possible argument taken from the initial dialog. A dialog is good if it is close to that ideal tree. Ideal dialogs have positive properties with respect to conciseness, namely they contain only useful and relevant arguments for the subject of the dialog. Moreover for every decisive move its argument belongs to all ideal trees.

From the results of this paper, it seems natural that a protocol generates dialogs of good quality if (1) irrelevant and not useful moves are penalized until there is a set of arguments that relate them to the subject (2) adding arguments in favor of the subject that are attacked by already present arguments has no interest (since they do not belong to any ideal tree). By doing so, the generated dialogs are more *concise* (*i.e.*, all the uttered arguments have an impact on the result of the dialog), and more *efficient* (*i.e.*, they are the minimal dialogs that can be built from the information exchanged and that reach the goal of the persuasion).

Note that in our proposal, the order of the arguments has not to be constrained since the generated graph does not take it into account. The only thing that matters in order to obtain a conclusion is the final set of interactions between the exchanged arguments. But the criteria of being relevant to the previous move or at least to a move not too far in the dialog sequence could be taken into account for analyzing dialog quality. Moreover, all the measures already defined in literature and cited in the introduction could also be used to refine the proposed preference relation on dialogs and finally could help to formalize general properties of protocols in order to generate good dialogs.

Furthermore, it may be the case that from the set of formulas involved in a set of arguments, new arguments may be built. This gives birth to a new set of arguments and to a new set of attack relations called complete argumentation system associated with a dialog. Hence, it could be interesting to define dialog trees on the basis of the complete argumentation system then more efficient dialogs could be obtained (but this is not guaranteed). However, some arguments of the complete argumentation system may require the cooperation of the agents. It would mean that in an ideal but practicable dialog, the order of the utterance of the arguments would be constrained by the fact that each agent should be able to build each argument at each step.

## References

[1] L. Amgoud and C. Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *Int. Journal of Automated Reasoning*, Volume 29 (2): 125-169, 2002.

[2] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34: 197–216, 2002.

[3] L. Amgoud and F. Dupin de Saint-Cyr. Measures for persuasion dialogs: a preliminary investigation. In *2nd Int. Conf. on Computational Models of Argument*, pages 13–24, 2008.

[4] L. Amgoud and F. Dupin de Saint-Cyr. Extracting the core of a persuasion dialog to evaluate its quality. In *10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2009)*, Verona, Italy, July 1–3, 2009.

[5] L. Amgoud and N. Maudet. Strategical considerations for argumentative agents (preliminary report). In *Proceedings of the 10th International Workshop on Non-Monotonic Reasoning NMR'2002 (Collocated with KR'2002), session Argument, Dialogue, Decision*, pages 409–417, 2002.

[6] L. Amgoud and N. Maudet and S. Parsons. Modelling dialogues using argumentation. In *Proc. of the International Conference on Multi-Agent Systems*, pages 31–38, Boston, MA, 2000.

[7] T. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *J. of Logic and Computation*, 13 (3): 429–448, 2003.

[8] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and $n$-person games. *Artificial Intelligence Journal*, 77: 321–357, 1995.

[9] P. E. Dunne and T. J. Bench-Capon. Two Party Immediate Response Disputes: Properties and Efficiency. *Artificial Intelligence*, 149: 221–250, 2003.

[10] T. F. Gordon. The pleadings game, *Artificial Intelligence and Law*, 2: 239–292, 1993.

[11] S. Parsons and P. McBurney. Games that agents play: A formal framework for dialogues between autonomous agents. *J. of Logic, Language, and Information*, 11 (3): 315–334, 2002.

[12] H. Prakken, Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15: 1009–1040, 2005.

[13] P. Torroni. A study on the termination of negotiation dialogues. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 1223–1230, ACM 2002.

[14] S. Zabala and I. Lara and H. Geffner. Beliefs, Reasons and Moves in a model for argumentation dialogues. In *Proc. 25th Latino-American Conf. on Computer Science*, 1999.

Leila Amgoud
IRIT – 118, route de Narbonne
31062 Toulouse Cedex 9, France
amgoud@irit.fr

Florence Dupin de Saint-Cyr
IRIT – 118, route de Narbonne
31062 Toulouse Cedex 9, France
bannay@irit.fr

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 23 (36) 2011

**Katarzyna Budzyńska**
Cardinal Stefan Wyszyński University in Warsaw

**Magdalena Kacprzak**
Bialystok University of Technology

# MODEL CHECKING OF PERSUASION IN MULTI-AGENT SYSTEMS

**Abstract**: The paper presents the method of model checking applied to verification of persuasive inter-agent communication. The model checker Perseus is designed on the basis of a logic of actions and graded beliefs $\mathcal{AG}_n$ introduced by Budzyńska and Kacprzak. The software tool makes it possible to semantically verify satisfaction of $\mathcal{AG}_n$ formulas which describe different properties of a multi-agent system in a given model, and to perform parametric verification that enables searching for answers to questions about these properties.
**Keywords**: model-checking, modal logic, multi-agent systems, persuasive arguments, dialogue games

## 1. Introduction

The common method of verification of multi-agent systems is model checking technique (see e.g. [9, 11, 15]). The paper presents how this method can be applied to examine the properties of inter-agent persuasive communication. A software tool designed to verify persuasion in multi-agent systems is called **Perseus** [6]. It is built upon the Logic of Actions and Graded Beliefs $\mathcal{AG}_n$ [2]. The Perseus model checker offers two main options of investigation. First, it can semantically *verify satisfaction of formulas* of the $\mathcal{AG}_n$ language which describe properties of persuasion in a given model. In this case, the tool performs the standard model checking. Second, it can *search for answers to questions* of three kinds – questions about the degrees of uncertainty, questions about the sequence of arguments that should be executed and questions about the agents participating in the process of persuasion. In this case, the tool uses the new method of parametric verification introduced by Budzyńska, Kacprzak and Rembelski [6].

The most typical kind of inter-agent persuasive communication is **persuasion dialogue** [16]. It is a dialogue of which initial situation is a conflict of opinion and the aim is to resolve this conflict and thereby influence the change of agents' beliefs or commitments (i.e. beliefs declared by an agent). This rises a general question about what impact on beliefs and commitments has a given persuasion. In consequence, we may ask about the degree and the scenario of belief and commitment changes, the factors that influence them, the strength of different types of arguments and their arrangements, the credibility of persuader, the strategies that allow the victory in a dialogue game etc.

Formal systems for dialogues are often built in a game-theoretic style, i.e. speech acts performed in a dialogue are treated as moves in a **dialogue game** and rules for their appropriateness are formulated as rules of the game (see [13] for an overview). In this paper, we present the application of verification methods to a persuasion dialogue system introduced by Prakken [12]. A *dialogue system* for argumentation is defined as a pair $(\mathcal{L}; \mathcal{D})$, where $\mathcal{L}$ is a logic for defeasible argumentation and $\mathcal{D}$ is a dialogue system proper. A *logic for defeasible argumentation* $\mathcal{L}$ is a tuple $(L_t, R, Args, \rightarrow)$, where $L_t$ (the topic language) is a logical language, $R$ is a set of inference rules over $L_t$, $Args$ (the arguments) is a set of AND-trees of which the nodes are in $L_t$ and the AND-links are inferences instantiating rules in $R$, and $\rightarrow$ is a binary relation of defeat defined on $Args$. For any argument $A$, $prem(A)$ is the set of leaves of $A$ (its premises) and $conc(A)$ is the root of $A$ (its conclusion).

A *dialogue system proper* is a triple $\mathcal{D} = (L_c; P; C)$ where $L_c$ (the communication language) is a set of locutions (utterances), $P$ is a protocol for $L_c$, and $C$ is a set of effect (commitment) rules of locutions in $L_c$, specifying the effects of the locutions on the participants' commitments. Agents may perform five types of **communication moves** (dialogue actions) in a dialogue game: $claim(\alpha)$ – the speaker asserts that $\alpha$ is the case, $why(\alpha)$ – the speaker challenges $\alpha$ and asks for reasons why it would be the case, $concede(\alpha)$ – the speaker admits that $\alpha$ is the case, $retract(\alpha)$ – the speaker declares that he is not committed (any more) to $\alpha$, $argue(A)$ – the speaker provides an argument $A$. Every utterance from $L_c$ can influence participants commitments. $C(d, i)$ denotes a player $i$'s **commitments** at a stage of a dialogue $d$. In a dialogue game, an agent may adopt a strategy to achieve a desired goal which could be to make its adversary become committed to the agent's claim (see e.g. [10], [14], [1]).[1]

The remainder of this paper is organized as follows. Section 2 presents the model checking technique. Section 3 shows the $\mathcal{AG}_n$ logic. Section 4

[1] For present purposes a more detailed definitions are not needed. For the full details the reader is referred to [12].

presents the model checker Perseus. Finally, Section 5 discusses the types of properties of persuasion that may be verified by the Perseus system.

## 2. The model checking technique

The commonly applied method which allows for semantic verification of multi-agent systems and thus their communication is *model checking*. Model checking is considered as a one of the most spectacular applications of computer science. Testing the correctness of a given software system under their correctness conditions (i.e. specification) is crucial task which must be solved before this system will find the application in commercial exploitation. Potential non detected errors in programs like decision support for air traffic control or quality control can have financial effects or could be a threat to health or life of people. Computer simulations and computations allow for avoiding very expensive and time-consuming experiments. The main problem which appears in automatic verification is state explosion problem. There are diverse methods for dealing with this problem. However, in practice, the most effective results bring application of symbolic methods based on satisfiability of propositional formulas.

Model checking is the decision problem that takes either a program $P$ or its more extensive representation $\mathcal{M}_{\mathcal{P}}$ (as a transition system), and a logical specification $\alpha$ which truth value should be determined. Then, the model checking problem for $P$ asks whether $\mathcal{M}_P, s \models \alpha$ for a given state $s$ of the model $\mathcal{M}_{\mathcal{P}}$. The model checking experiments executed by Walton have focused on the termination of Multi-Agent Protocol (in particular, auction protocols) [15]. The key difference between this approach and our research is that we want to verify the behavior of a whole multi-agent system, and not the properties of a language. We focus on the question what effects the dialogue actions can bring for a system, rather than the questions such as e.g. if an agent sends a sincere message. For instance, we want to ask what is the most effective (e.g. the shortest) sequence of speech acts that enables an agent $i$ to achieve his goal (e.g. to persuade his opponent $\bar{i}$ to accept $i$'s topic $t$).

For verification of dialogue games we use the model checker called Perseus [7]. Verification of epistemic and temporal formulas can be performed by other software tools like VerICS [9] or MCMAS [11], however, the important advantage of the Perseus system is that it performs not only pure model checking, but also parametric verification. It means that, given a model of a system, it can test automatically whether this model meets a given specification or given an input expressions with unknowns it can determine for which values of these unknowns the obtained logical formula is true in this model. Furthermore, Perseus is not limited to verification of formulas with epistemic and dynamic operators, but is already designed and adjusted to analyze phenomena related to agents persuasive actions with the use of graded doxastic modalities and with probabilistic modalities.

## 3. Logic of actions and graded beliefs $\mathcal{AG}_n$

In this section, we present the Logic of Actions and Graded Beliefs $\mathcal{AG}_n$ [2] extended with the components needed for representation of Prakken's dialogue system [3] and probabilistic beliefs [7].

### 3.1. Formal syntax of the language

Let $Agt = \{1, \ldots, n\}$ be a set of names of *agents*, $V_0$ be a set of *propositional variables*, $\Pi_0^{ph}$ a set of *physical actions*, and $\Pi_0^v$ a set of *verbal actions*. Further, let ; denote a programme connective which is a sequential composition operator. It enables to compose *schemes of programs* defined as finite sequences of atomic actions: $a_1; \ldots; a_k$. Intuitively, the program $a_1; a_2$ for $a_1, a_2 \in \Pi_0^{ph}$ means "Do $a_1$, then do $a_2$". The set of all schemes of physical programs we denote by $\Pi^{ph}$. In similar way, we define a set $\Pi^v$ of schemes of programs constructed over $\Pi_0^v$. The set of $F$ all *well-formed expressions* of the extended $\mathcal{AG}_n$ is given by the following Backus-Naur form:

$$\alpha ::= p|\neg\alpha|\alpha \vee \alpha|M_i^k\alpha|\mathbf{P}_i(\alpha) \geq q|\Diamond(i:P)\alpha|\mathbf{C}_i\alpha|$$

$$\langle i\rangle \mathrm{G}\alpha|\langle i\rangle^q \mathrm{G}\alpha|\langle\!\langle i\rangle\!\rangle \mathrm{G}\alpha|\langle i\rangle_f \mathrm{G}\alpha|\langle i\rangle_f^q \mathrm{G}\alpha|\langle\!\langle i\rangle\!\rangle_f \mathrm{G}\alpha|\langle\!\langle i\rangle\!\rangle \mathrm{X}\alpha|\langle\!\langle i\rangle\!\rangle \alpha \mathrm{U}\beta,$$

where $p \in V_0$, $k \in \mathbb{N}$, $q \in [0,1]$, $i \in Agt$, $P \in \Pi^{ph}$ or $P \in \Pi^v$ and $f$ is a strategy.

We use also the following abbreviations:

- $B_i^k\alpha$ for $\neg M_i^k\neg\alpha$,
- $M!_i^k\alpha$ where $M!_i^0\alpha \Leftrightarrow \neg M_i^0\alpha$, $M!_i^k\alpha \Leftrightarrow M_i^{k-1}\alpha \wedge \neg M_i^k\alpha$, if $k > 0$,
- $M!_i^{k_1,k_2}\alpha$ for $M!_i^{k_1}\alpha \wedge M!_i^{k_2}(\alpha \vee \neg\alpha)$,
- $\Box(i:P)\alpha$ for $\neg\Diamond(i:P)\neg\alpha$.
- $\mathbf{P}_i(\omega) > q$, $\mathbf{P}_i(\omega) = q$, $\mathbf{P}_i(\omega) < q$, $\mathbf{P}_i(\omega) \leq q$ defined from $\mathbf{P}_i(\omega) \geq q$ in the classical way.

### 3.2. Some intuitions

Intuitions concerning the most frequently used $\mathcal{AG}_n$ formulas is described below. The **belief formula** $M!_i^{k_1,k_2}\alpha$ says that an agent $i$ considers $k_2$ doxastic alternatives and in $k_1$ of them $\alpha$ holds. In other words, the agent $i$ believes $\alpha$ with degree $\frac{k_1}{k_2}$. The **probability formula** $\mathbf{P}_i(\alpha) \geq q$ informally says that the agent $i$ believes with probability higher or equal to $q$ that $\alpha$ holds. The **commitment formula** $\mathbf{C}_i\alpha$ says that a claim $\alpha$ is a commitment of the agent $i$. The $\mathcal{AG}_n$ language contains also **dynamic formulas** which allow the representation of physical actions which modify states of a model, and verbal actions which change a whole model. It means that physical actions influence agents' environment while verbal actions influence their perception of the environment. The formula $\Diamond(i : P)\alpha$ says that after executing a sequence of (persuasion) actions $P$ by the agent $i$, condition $\alpha$ may hold. Finally we explain the meaning of **strategic formulas**. $\langle i\rangle \mathrm{G}\alpha$ says that there *exists* a strategy of $i$ and there *exists* a computation consistent with this strategy such that in all states of this computation $\alpha$ is true. The formula $\langle i\rangle^q \mathrm{G}\alpha$ expresses that agent $i$ has such a strategy with degree of success which is higher than $q$. $\langle\langle i\rangle\rangle \mathrm{G}\alpha$ expresses that there *exists* such a strategy which *always* leads to success regardless of the other agents' actions, i.e. the agent $i$ has a *winning strategy.* The operator $\langle i\rangle_f \mathrm{G}\alpha$ says that for the strategy $f$ there *exists* a computation consistent with this strategy such that in all states of this computation $\alpha$ is true. The operator $\langle i\rangle_f^q \mathrm{G}\alpha$ expresses that the strategy $f$ of agent $i$ has degree higher or equal to $q$. The last operator we use is $\langle\langle i\rangle\rangle_f \mathrm{G}\alpha$ expresses that the strategy $f$ is a winning strategy.

### 3.3. Kripke model

All $\mathcal{AG}_n$ formulas are interpreted over the semantic model which is an extended Kripke structure.

**Definition 1**

By a *semantic model* we mean a Kripke structure

$$\mathrm{M} = (Agt, S, RB, I^{ph}, P, C, v)$$

where

- $Agt$ is a set of agents' names,
- $S$ is a non-empty set of states (the universe of the structure),
- $RB : Agt \longrightarrow 2^{S\times S}$ is a doxastic function which assigns to every agent a binary relation,

- $I^{ph} : \Pi_0^{ph} \longrightarrow (Agt \longrightarrow 2^{S \times S})$ is an interpretation of physical actions,
- $P : Agt \to (S \times S \to [0,1])$ is a probability (partial) function defined for every $i \in Agt$ and $(s, s') \in RB(i)$ such that for every agent $i \in Agt$ and $s \in S$, $\sum_{\{s' : (s,s') \in RB(i)\}} P(i)(s, s') = 1$,
- $C : S \times Agt \longrightarrow 2^F$ is a commitment function,
- $v$ is a valuation function, $v : S \longrightarrow \{\mathbf{0}, \mathbf{1}\}^{V_0}$.

Function $I^{ph}$ can be extended in a simple way to define interpretation of any program scheme. Let $I^{ph}_{\Pi^{ph}} : \Pi^{ph} \longrightarrow (Agt \longrightarrow 2^{S \times S})$ be a function such that $I^{ph}_{\Pi^{ph}}(P_1; P_2)(i) = I^{ph}_{\Pi^{ph}}(P_1)(i) \circ I^{ph}_{\Pi^{ph}}(P_2)(i) = \{(s, s') \in S \times S : \exists_{s'' \in S} \ ((s, s'') \in I^{ph}_{\Pi^{ph}}(P_1)(i) \text{ and } (s'', s') \in I^{ph}_{\Pi^{ph}}(P_2)(i))\}$ for $P_1, P_2 \in \Pi^{ph}$ and $i \in Agt$.

Further, we define a function $I^v$ which is an interpretation for verbal actions.

**Definition 2**

Let $\mathcal{CM}$ be a class of models and $\mathcal{CMS}$ be a set of pairs $(\mathrm{M}, s)$ where $\mathrm{M} \in \mathcal{CM}$ and $s$ is a state of the model M. An interpretation for verbal actions $I^v$ is a function:

$$I^v : \Pi_0^v \longrightarrow (Agt \longrightarrow 2^{\mathcal{CMS} \times \mathcal{CMS}}).$$

We allow different verbal actions to be executed during persuasion process. Therefore, no restrictions on $I^v$ are assumed in the general definition. An interpretation for verbal actions will obtain different specifications depending on the type of actions and the applications of the formal model. Moreover, verbal actions do not have to convey a true information. This is particularly important, if we want to use the formal framework to represent persuasion. Agents may try (successfully or not) to influence others using false messages (since they are insincere or have incomplete knowledge). Thus, we assume that $I^v$ does not depend on the truth or falsity conditions of the announced formula. Interpretation $I^v_{\Pi^v}$ of all verbal programs is defined similarly to the function $I^{ph}_{\Pi^{ph}}$.

Before the semantics of several kinds of strategy modalities will be defined, we need to formalize the notion of a *strategy*. Let

$$\delta : \mathcal{CMS} \times Agt \to 2^{(\Pi^{Ph} \cup \Pi^v)}$$

be a function mapping a triple consisting of a model, a state of this model and an agent to a set of actions. These actions are assumed to be actions which the agent can perform next. In fact this function determines *transition* *function*, i.e. indicates models and states of these models reachable from a given state of a given model by a given agent.

**Definition 3**

A **computation** is a sequence

$$(\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), (\mathrm{M}_2, s_2), \ldots$$

such that for every $k \geq 0$, there exists an action $a_k$ and an agent $i_k$ such that $a_k \in \delta(\mathrm{M}_k, s_k, i_k)$ and $((\mathrm{M}_k, s_k), (\mathrm{M}_{k+1}, s_{k+1})) \in I(a_k, i_k)$ where $I$ is the interpretation of action $a_k$, i.e., $I = I^{ph}$ if $a_k$ is a physical action and $I = I^v$ if $a_k$ is a verbal action.

Intuitively by a computation we mean a sequence of pairs $(\mathrm{M}_k, s_k)$, a model and a state of this model, such that for every position $k$, $(\mathrm{M}_{k+1}, s_{k+1})$ is a result of performing an action $a_k$ by an agent $i_k$ at the state $s_k$ of the model $\mathrm{M}_k$.

**Definition 4**

By a **strategy** for an agent $i$ we call a mapping $f_i : M^{<\infty} \to 2^M$ which assigns to every finite dialogue $d = m_0, m_1, \ldots, m_k \in M^{<\infty}$ in which it is $i$'s turn, i.e., $i \in T(d)$, a move $m \in M$ such that $m \in Pr(d)$.

In other words, a strategy function returns a move which is allowed by the protocol $P$ after a dialogue $d$ where $i$ is to move. We say that a dialogue $d = m_0, m_1, \ldots$ is consistent with a strategy $f_i$ iff for every $k \geq 1$ if $i = pl(m_k)$ then $m_k \in f_i(m_0, \ldots, m_{k-1})$ and for $k = 0$ if $i = pl(m_k)$ then $m_k \in f_i(\emptyset)$, i.e., every move of agent $i$ is determined by the function $f_i$.

Next, we define the ***outcomes*** of $f_i$, i.e., a set of computations which are consistent with this strategy. Let $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), (\mathrm{M}_2, s_2), \ldots$ be a computation, then

$$\lambda \in out((\mathrm{M}, s), f_i) \quad \text{iff} \quad (\mathrm{M}_0, s_0) = (\mathrm{M}, s) \quad \text{and}$$

there exists a dialogue $d = m_0, m_1, \ldots$ consistent with $f_i$ such that

$$\text{for every } k \geq 0, \quad s(m_k) \in \delta(\mathrm{M}_k, s_k, pl(m_k))$$

and

$$((\mathrm{M}_k, s_k), (\mathrm{M}_{k+1}, s_{k+1})) \in I(pl(m_k), s(m_k)).$$

Intuitively, a computation is consistent with a strategy if it is determined by a dialogue consistent with the strategy.

### 3.4. Interpretation of formulas

The *semantics* of formulas of the $\mathcal{AG}_n$ logic is defined with respect to a model M, i.e., for a given structure $\mathrm{M} = (S, RB, I^{ph}, P, C, v)$ and a given state

$\mathrm{M}, s \models p$ iff $v(s)(p) = \mathbf{1}$, for $p \in V_0$,

$\mathrm{M}, s \models \neg\alpha$ iff $\mathrm{M}, s \not\models \alpha$,

$\mathrm{M}, s \models \alpha \vee \beta$ iff $\mathrm{M}, s \models \alpha$ or $\mathrm{M}, s \models \beta$,

$\mathrm{M}, s \models M_i^k \alpha$ iff $|\{s' \in S : (s, s') \in RB(i)$ and $\mathrm{M}, s' \models \alpha\}| > k$, $k \in \mathbb{N}$,

$\mathrm{M}, s \models \Diamond(i : P)\alpha$ iff $\exists_{s' \in S}$ $((s, s') \in I^{ph}_{\Pi^{ph}}(P)(i)$ and $\mathrm{M}, s' \models \alpha)$ for $P \in \Pi^{ph}$ or $\exists_{(\mathrm{M}', s') \in \mathcal{CMS}}$ $(((\mathrm{M}, s), (\mathrm{M}', s')) \in I^v_{\Pi^v}(P)(i)$ and $\mathrm{M}', s' \models \alpha)$ for $P \in \Pi^v$,

$\mathrm{M}, s \models \mathbf{P}_i(\alpha) \geq q$ iff $\sum_{\{s' \in S | (s,s') \in RB(i) \text{ and } \mathrm{M}, s' \models \alpha\}} P(i)(s, s') \geq q$,

$\mathrm{M}, s \models \mathbf{C}_i \alpha$ iff $\alpha \in C(s, i)$,

$\mathrm{M}, s \models \langle i \rangle \mathrm{G}\alpha$ iff there exists a strategy $f_i$ such that for some computation $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_0, s_0), \ldots \in out((\mathrm{M}, s), f_i)$, and for all positions $k \geq 0$, we have $(M_k, s_k) \models \alpha$,

$\mathrm{M}, s \models \langle i \rangle^q \mathrm{G}\alpha$ iff there exists a strategy $f_i$ such that $\frac{k_1}{k_2} \geq q$ for $k_2$ being the number of all computations $\lambda \in out((\mathrm{M}, s), f_i)$ and $k_1$ being the number of computations $\lambda \in out((\mathrm{M}, s), f_i)$ in which every state satisfy $\alpha$,

$\mathrm{M}, s \models \langle\langle i \rangle\rangle \mathrm{G}\alpha$ iff there exists a strategy $f_i$ such that for all computations $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), \ldots \in out((M, s), f_i)$, and for all positions $k \geq 0$, we have $\mathrm{M}_k, s_k \models \alpha$,

$\mathrm{M}, s \models \langle i \rangle_f \mathrm{G}\alpha$ iff for some computation $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), \ldots \in out((M, s), f)$, and for all positions $k \geq 0$, we have $\mathrm{M}_k, s_k \models \alpha$,

$\mathrm{M}, s \models \langle i \rangle_f^q \mathrm{G}\alpha$ iff $f$ is a strategy such that $\frac{k_1}{k_2} \geq q$ for $k_2$ being the number of all computations $\lambda \in out((\mathrm{M}, s), f)$ and $k_1$ being the number of computations $\lambda \in out((\mathrm{M}, s), f)$ in which every state satisfy $\alpha$,

$\mathrm{M}, s \models \langle\langle i \rangle\rangle_f \mathrm{G}\alpha$ iff for all computations $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), \ldots \in out((M, s), f)$, and for all positions $k \geq 0$, we have $\mathrm{M}_k, s_k \models \alpha$,

$\mathrm{M}, s \models \langle\langle i \rangle\rangle \mathrm{X}\alpha$ iff there exists a strategy $f_i$ such that for all computations $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), \ldots \in out((M, s), f_i)$, we have $\mathrm{M}_1, s_1 \models \alpha$,

$\mathrm{M}, s \models \langle\langle i \rangle\rangle \alpha \mathrm{U} \beta$ iff there exists a strategy $f_i$ such that for all computations $\lambda = (\mathrm{M}_0, s_0), (\mathrm{M}_1, s_1), \ldots \in out(s, f_i)$, there exists a position $k \geq 0$ such that $\mathrm{M}_k, s_k \models \beta$ and for all positions $0 \leq j < k$, we have $\mathrm{M}_j, s_j \models \alpha$.

Based on commitment rules defined in [12] we introduce a specification of dialogue actions. Since speech acts are verbal actions they move a system from a model M to a new model M$'$. For instance, an action $claim(\alpha)$ performed at a state $s$ of a model M moves a multi-agent system to a model M$'$ in which a new commitment function $C'$ is defined in such a way that the new set of commitments of the performer of the action at $s$ equals to the old one enriched with $\alpha$. Notice that *claim* and *concede* are two different speech acts and are used with two different intentions. An agent "*claims* $\alpha$" when he publicly announces that he is committed to $\alpha$. Whereas, an agent "*concedes* $\alpha$" when he agrees with his opponent that $\alpha$ holds. Therefore, it is a reply to an opponent's argumentation for $\alpha$. Nevertheless, formal specification of these actions is exactly the same. Formally, they both add a formula $\alpha$ to the set of commitments of the performer. A formal definition of interpretation of dialogue actions $I^v$ is as follows:

1. **claim**:
   $((\mathrm{M}, s), (\mathrm{M}', s)) \in I^v(claim(\alpha))(i)$ iff $\mathrm{M}' = (S, RB, I^{ph}, v, C')$
   where
   - $C'(s,i) = C(s,i) \cup \{\alpha\}$ and
   - $C'(s',i') = C(s',i')$ for $s' \neq s$ or $i' \neq i$,

2. **concede**:
   $((\mathrm{M}, s), (\mathrm{M}', s)) \in I^v(concede(\alpha))(i)$ iff $\mathrm{M}' = (S, RB, I^{ph}, v, C')$
   where
   - $C'(s,i) = C(s,i) \cup \{\alpha\}$ and
   - $C'(s',i') = C(s',i')$ for $s' \neq s$ or $i' \neq i$,

3. **retract**:
   $((\mathrm{M}, s), (\mathrm{M}', s)) \in I^v(retract(\alpha))(i)$ iff $\mathrm{M}' = (S, RB, I^{ph}, v, C')$
   where
   - $C'(s,i) = C(s,i) \backslash \{\alpha\}$ and
   - $C'(s',i') = C(s',i')$ for $s' \neq s$ or $i' \neq i$,

4. **argue**:
   $((\mathrm{M}, s), (\mathrm{M}', s)) \in I^v(argue(A))(i)$ iff $\mathrm{M}' = (S, RB, I^{ph}, v, C')$
   where
   - $C'(s,i) = C(s,i) \cup prem(A) \cup conc(A)$
   - $C'(s',i') = C(s',i')$ for $s' \neq s$ or $i' \neq i$,

5. **why**:
   $((\mathrm{M}, s), (\mathrm{M}', s)) \in I^v(why(\alpha))(i)$ iff $\mathrm{M}' = \mathrm{M}$.

## 4. Model checker Perseus

The Perseus system is a software tool designed for an automatic many-sided analysis of persuasive multi-agent systems. It was designed in 2008 by Budzyńska, Kacprzak and Rembelski and is still developed [6]. Its aim is to analyze persuasion ability of multi-agent systems given their formal model. Until now, Perseus can deal with features concerning graded beliefs of agents, probabilistic beliefs of agents and the impact of persuasive actions on agents' beliefs and activities.

Given a semantic model $M$ of a system, the task of the Perseus system is to automatically analyze its properties. It could be done twofold: by model checking or by parametric verification. Application of **model checking** method allows for testing whether a $\mathcal{AG}_n$ formula is true in a given state of the model $M$. In other words, using model checking technique Perseus tests whether some specific property holds in a multi-agent system represented by the model $M$. **Parametric verification** was introduced by the authors of the Perseus system. This method allows Perseus to look for answers to questions about diverse properties of systems under consideration and, in consequence, allows to analyze these systems in an automatic way. In particular, questions can concern

- agents – is there an agent who can influence somebody's beliefs?, who can do it?, who can achieve a success?
- beliefs and degrees of beliefs – does an agent believe a claim?, what is a degree of his uncertainty about this claim?
- results of actions – whether a degree of agent's belief can change after execution of a given action or sequence of actions?, which actions should be executed in order to convince an agent that a claim is true?

The system input data of the Perseus tool, i.e. the **input question**, is a triple $(M, s, \phi)$, where $M$ is a model described by an arbitrary specification of a model (see [6]), $s$ is a state of the model $M$ and $\phi$ is the **input expression**. The input expression is defined by the following BNF:

$$\phi ::= \omega|\neg\phi|\phi \vee \phi|M_i^d\phi|\Diamond(i : P)\phi|M_i^?\omega|\Diamond\,(i :?)\,\omega|\mathbf{P}_i\,(\omega) \geq ?|$$

$$M_?^d\omega|\Diamond\,(? : P)\,\omega|\mathbf{P}_?\,(\omega) \geq q,$$

where $\omega ::= p|\neg\omega|\omega \vee \omega|M_i^d\omega|\Diamond\,(i : P)\,\omega|\mathbf{P}_i(\omega) \geq q$ and $p \in V_0$, $d \in \mathbb{N}$, $P \in \Pi^{ph}$ or $P \in \Pi^v$, $i \in Agt$ as well as $q \in [0; 1]$. Therefore the language of extended $\mathcal{AG}_n$ logic is a sublanguage of the Perseus system input expressions (what follows is that other modalities $B_i^d\omega$, $M!_i^d\omega$, $M!_i^{d_1,d_2}\omega$, $\Box\,(i : P)\,\omega$,

$\mathbf{P}_i(\omega) > q$, $\mathbf{P}_i(\omega) = q$, $\mathbf{P}_i(\omega) < q$, $\mathbf{P}_i(\omega) \leq q$, can be derived in the standard way).

Perseus system accepts two types of the input expressions:

– **unknown free expressions**, where grammar productions

$$M_i^? \omega | \Diamond (i :?) \omega | \mathbf{P}_i(\omega) \geq ? | M_?^d \omega | \Diamond (? : P) \omega | \mathbf{P}_? (\omega) \geq q$$

are not allowed,

– **one-unknown expression**, where only one of the grammar productions

$$M_i^? \omega | \Diamond (i :?) \omega | \mathbf{P}_i(\omega) \geq ? | M_?^d \omega | \Diamond (? : P) \omega | \mathbf{P}_? (\omega) \geq q$$

is allowed.

Next the Perseus system executes a parametric verification of an input question, i.e. tests if (both unknown free and one-unknown expressions) and when (only one-unknown expressions) the expression $\phi$ becomes a formula of the extended $\mathcal{AG}_n$ logic $\phi^*$ such that $M, s \models \phi^*$.

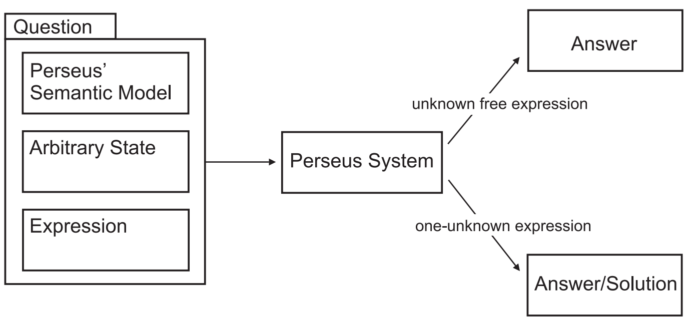



**Figure 1. The idea of the Perseus system**

In case of unknown free expressions we have $\phi^* = \phi$, i.e a standard model verification is done. In the other case a formula $\phi^*$ is obtained from $\phi$ by swapping all ? symbols for appropriate values either from set $\{0, 1, \ldots |S|\}$ or $Agt$ or $\Pi^{ph}$ or $\Pi^v$ or $[0; 1]$. Finally the system output data, i.e the **output answer**, is given. The output answer is *true* if $M, s \models \phi^*$ and *false* otherwise (see Fig. 1). As soon as the output answer is determined, the **solution set** $X$ for the one-unknown expression is presented, where:

– $X \subseteq \{0, 1, \ldots |S|\}$, for an expression $\phi$ with one unknown of type $M_i^? \omega$, $B_i^? \omega$, $M!_i^? \omega$, $M!_i^{?,d_2} \omega$, $M!_i^{d_1,?} \omega$,

– $X \subseteq \{0, 1, \ldots |S|\} \times \{0, 1, \ldots |S|\}$, for an expression $\phi$ with one unknown of type $M!_i^{?_1,?_2} \omega$,

– $X \subseteq Agt$, for an expression $\phi$ with one unknown of type $M_?^d \omega$, $B_?^d \omega$, $M!_?^d \omega$, $M!_?^{d_1,d_2} \omega$, $\Diamond (? : P)\, \omega$, $\Box (? : P)\, \omega$, $\mathbf{P}_? (\omega) \geq q$, $\mathbf{P}_? (\omega) > q$, $\mathbf{P}_? (\omega) = q$, $\mathbf{P}_? (\omega) < q$, $\mathbf{P}_? (\omega) \leq q$,

– $X \subseteq \Pi^{ph}$ or $X \subseteq \Pi^v$, for an expression $\phi$ with one unknown of type$\Diamond (i :?)\, \omega$, $\Box (i :?)\, \omega$,

– $X \subseteq [0; 1]$, for an expression $\phi$ with one unknown of type $\mathbf{P}_i (\omega) \geq ?$, $\mathbf{P}_i (\omega) > ?$, $\mathbf{P}_i (\omega) = ?$, $\mathbf{P}_i (\omega) < ?$, $\mathbf{P}_i (\omega) \leq ?$.

In order to find an answer to the input question $(M, s, \phi)$, the Perseus system executes the syntax analysis of the expression $\phi$. The analysis is based on the standard descent recursive method. As a result a syntax tree of expression $\phi$ is created. All inner nodes of such a tree represent either Boolean operators or $\mathcal{AG}_n$ logic modalities while all outer nodes stand for either propositional variables or unknown. The solution for an arbitrary unknown is reached in the following way:

– if an unknown type is $M_i^? \omega$, $B_i^? \omega$, $M!_i^? \omega$, $M!_i^{?,d_2} \omega$, $M!_i^{d_1,?} \omega$, $M!_i^{?_1,?_2} \omega$, then the counting method is applied, i.e. all states, which are reachable via a doxastic relation of the agent $i$, and in which the claim $\omega$ is satisfied or refuted respectively, are counted,

– if an unknown type is $M_?^d \omega$, $B_?^d \omega$, $M!_?^d \omega$, $M!_?^{d_1,d_2} \omega$, $\Diamond (? : P)\, \omega$, $\Box (? : P)\, \omega$, $\mathbf{P}_? (\omega) \geq q$, $\mathbf{P}_? (\omega) > q$, $\mathbf{P}_? (\omega) = q$, $\mathbf{P}_? (\omega) < q$, $\mathbf{P}_? (\omega) \leq q$, say $\mathbf{P}_? (\omega) \geq q$, then for every agent $i \in Agt$ the property $\mathcal{M}, s \models \mathbf{P}_i (\omega) \geq q$ is tested,

– if an unknown type is $\Diamond (i :?)\, \omega$, $\Box (i :?)\, \omega$, then a nondeterministic finite automaton, which represents all possible argumentation $P \in \Pi$ such that respectively $\mathcal{M}, s \models \Diamond (i : P)\, \omega$ or $\mathcal{M}, s \models \Box (i : P)\, \omega$ holds, is created,

– if an unknown type is $\mathbf{P}_i (\omega) \geq ?$, $\mathbf{P}_i (\omega) > ?$, $\mathbf{P}_i (\omega) = ?$, $\mathbf{P}_i (\omega) < ?$, $\mathbf{P}_i (\omega) \leq ?$, then the **summing method** is applied, i.e. probabilistic coefficients of all states, which are reachable via doxastic relation of the agent $i$, and in which the claim $\omega$ is satisfied or refuted respectively, are add up.

If an unknown is a nested type, i.e. it is a part of claim of the extended $\mathcal{AG}_n$ logic operator, then its **solution set is bounded** by the outer modality/modalities. For example, if we consider an input question

$$\left(M, s, \Box\,(i:P)\, M!_j^1 \mathbf{P}_i\,(\omega) <?\right),$$

then the solution of the unknown $\mathbf{P}_i\,(\omega) <?$ is reduced firstly by the operator $M!$ and secondly by the operator $\Box$.

## 5. The properties of persuasion in multi-agent systems

In this section, we present the important properties of persuasion in multi-agent systems which could be examined with the use of a model-checker.

### 5.1. Influence on degrees of beliefs

In order to formally verify the properties of persuasion in multi-agent systems, Perseus searches for answers to **questions** expressed in the $\mathcal{AG}_n$ language. The first group of questions ask about the properties of persuasion related to influencing agents uncertainty. In our model, uncertainty is represented by two types of operators: graded and probabilistic modalities. Each of them encodes slightly different information. The graded belief formula $M!_{John}^{3,5}p$ expresses that there are 5 John's doxastic alternatives and in 3 of them $p$ holds, while the probabilistic formula $\mathbf{P}_{John}(p) = 0.6$ does not describe local properties of the model with such details, since equally John could allow 50 doxastic alternatives and in 30 of them $p$ would hold. Thus, in the latter case we are dealing with a loss of the information. In other words, a probabilistic formula says what is the uncertainty of an agent about a claim, but does not give any reasons. On the other hand, the probabilistic operator allows the verification of questions in which such a detailed information is not needed, but instead we are interested in all cases when an agent is uncertain in a specific degree. For example, we may ask if it is possible that after a persuasion John will believe a claim with the degree 0.6 regardless of how many doxastic alternatives he allows (i.e. no matter if there are 5 John's doxastic alternatives and in 3 of them $p$ holds or there are 50 John's doxastic alternatives and in 30 of them $p$ holds, and so on).

Perseus can check the property of influencing agent uncertainty with respect to unknown-free expressions using the standard model-checking technique, e.g. the tool can check whether:

- $M!_i^3\omega$, exactly 3 doxastic alternatives of the agent $i$ satisfy the claim $\omega$,
- $M_i^4 B_j^2\omega$, in more than 4 doxastic alternatives of the agent $i$ it is true that in at most 2 doxastic possibilities of the agent $j$ the claim $\omega$ is refuted,
- $\Diamond(i : P)\,\omega$, the execution of the argumentation $P$ by the agent $i$ may cause that the claim $\omega$ is satisfied,
- $\Box(i : P)\, M!_j^{2,4}\omega$, the execution of the argumentation $P$ by the agent $i$ can not cause that it is not true that in exactly 2 doxastic alternatives of the agent $j$ among exactly 4 his doxastic possibilities the claim $\omega$ is satisfied.

The Perseus tool can also check the property of influencing agent uncertainty with respect to one-unknown expressions using the parametric verification technique, e.g. it can check whether:

- $M_i^?\omega$, in more than how many doxastic alternatives of the agent $i$ the claim $\omega$ is satisfied?
- $M_?^d\omega$, for which agent is it true that in more than $d$ of his doxastic alternatives the claim $\omega$ is satisfied?
- $B_i^?\omega$, in at the most how many doxastic alternatives of the agent $i$ the claim $\omega$ is refuted?
- $B_?^d\omega$, for which agent is it true that in at most $d$ of his doxastic alternatives the claim $\omega$ is refuted?
- $M!_i^?\omega$, in exactly how many doxastic alternatives of the agent $i$ the claim $\omega$ is satisfied?
- $M!_?^d\omega$, for which agent is it true that in exactly $d$ of his doxastic alternatives the claim $\omega$ is satisfied?
- $M!_i^{?,d_2}\omega$, in exactly how many doxastic alternatives of the agent $i$, from exactly $d_2$ of his doxastic possibilities, the claim $\omega$ is satisfied?
- $M!_i^{d_1,?}\omega$, what is an exact number of all doxastic alternatives of the agent $i$, where in exactly $d_1$ of them the claim $\omega$ is satisfied?
- $M!_i^{?_1,?_2}\omega$, what is an exact number of all doxastic alternatives of the agent $i$ and in exactly how many of them the claim $\omega$ is satisfied?
- $M!_?^{d_1,d_2}\omega$, for which agent is it true that in exactly $d_1$ doxastic alternatives among exactly $d_2$ of his doxastic possibilities the claim $\omega$ is satisfied?
- $\Diamond(i :?)\,\omega$, for what argumentation is it true that its execution by the agent $i$ may cause that the claim $\omega$ is satisfied?

- $\Diamond\,(?:P)\,\omega$, for which agent is it true that his execution of the argumentation $P$ may cause the claim $\omega$ is satisfied?
- $\Box\,(i:?)\,\omega$, for what argumentation is it true that its execution by the agent $i$ can not cause that the claim $\omega$ is refuted?
- $\Box\,(?:P)\,\omega$, for which agent is it true that his execution of the argumentation $P$ can not cause that the claim $\omega$ is refuted?

In particular, the expression $M_i^{?_1,?_2}(\Diamond(i:a)M!_j^{10,10}\alpha)$ asks "What is a degree of an agent $i$'s belief that after using an argument $a$ an agent $j$ will believe $\alpha$ with a degree $\frac{10}{10}$?". The other question $\Diamond(i:a)M_j^{?_1,?_2}\alpha$ means "What will be a degree of an agent $j$'s belief about $\alpha$ after an argumentation $a$ performed by an agent $i$ ?" and the question $\Diamond(i:?)M_j^{10,10}\alpha$ means "What argumentation should an agent $i$ use to convince an agent $j$ to believe about $\alpha$ with a degree $\frac{10}{10}$?".

### 5.2. Conflict of opinion and persuasiveness

The initial condition for persuasion is a **conflict of opinion** [16]. Consider two agents which task is to increase the environment's temperature as soon as it drops below $10^0$C. Assume that one agent is not able to carry out the task on his own. It is possible only if agents cooperate. In our scenario a conflict will start when one of the agents believes that the temperature is lower than $10^0$C while the other does not – possibly because agents use different sources of information and thereby derive different conclusions. Observe that a conflict may appear not only when a proponent is absolutely sure about the claim and an audience is absolutely against it. It can also arise from the fact that the degrees of agents' beliefs differ or belong to different intervals. Say that degrees from $(\frac{1}{2}, 1]$ mean accepting the claim and degrees from $[0, \frac{1}{2}]$ mean rejecting the claim. Then, Perseus can verify the property with respect to the conflict of opinion checking if e.g. such a $\mathcal{AG}_n$ formula holds in a given model:

$$M!_{prop}^{3,4}(p_{t<10}) \wedge M!_{aud}^{1,4}(p_{t<10})$$

where *prop* and *aud* mean the proponent and the audience, respectively, and $p_{t<10}$ is a propositional variable which expresses that the temperature is lower than $10^0$C. This formula should be read as follows: "The proponent believes that the temperature is lower than $10^0$C with the degree $\frac{3}{4}$ and the audience believes the temperature is lower than $10^0$C with the degree $\frac{1}{4}$".

If the inter-agent conflict of opinion is resolved, we talk about the **success of persuasion**. The persuasion $P = (a_1; a_2; \ldots; a_k)$ can be successful when after performing actions $a_1$, $a_2$, …, $a_k$ by the proponent, it is possible that the audience will believe the claim with some expected degree. We assumed that an agent accepts the claim if he believes it in a degree higher than $\frac{1}{2}$. Then, the Perseus software can check if the proponent's persuasion may be successful:

$$\Diamond(prop : P)(M!^{3,4}_{aud}(p_{t<10})).$$

This formula states "If the proponent performs arguments $P$ then it is possible that the audience will believe the claim with the degree $\frac{3}{4}$".

In some circumstances, an agent may have a chance to achieve only the **subjective success**. That is, after execution of $P$ the proponent may believe that he achieved a goal while he actually did not, which means that he wrongly evaluated the results of his persuasion. Perseus may provide an answer whether there is a risk of such a situation:

$$\Diamond(prop : P)[M!^{4,4}_{prop}(M!^{3,4}_{aud}(p_{t<10})) \wedge \neg M!^{3,4}_{aud}(p_{t<10})].$$

The formula states "If the proponent performs arguments $P$ then it is possible he will believe that the audience is convinced with the degree $\frac{3}{4}$, but the audience will not believe the claim with this degree".

The other property that the model-checker can verify is whether the proponent **predicts** (believes) that he is able to succeed. Otherwise, he may not start persuasion even though he had all necessary means to win. Such a situation can be expressed by the formula:

$$M!^{0,4}_{prop}[\Diamond(prop : P)(M!^{4,4}_{aud}(p_{t<10}))] \wedge \Box(prop : P)(M!^{4,4}_{aud}(p_{t<10})).$$

If the formula holds in a given model, then the proponent is absolutely sure that his persuasion $P$ will fail (the proponent believes with the degree $\frac{0}{4}$ that the audience may become convinced to the claim with the degree $\frac{4}{4}$), while $P$ would actually lead him to success (after persuasion $P$ the audience will believe the claim with the desired degree $\frac{4}{4}$).

The **persuasiveness** depends on the arguments (their quality as well as length and order of argument sequence) and on the credibility of proponent. The same proponent convincing the same audience with the use of **different quality arguments** may arrive at different results of the persuasion. Say that if the proponent gives verbal argument "One of your thermometers is placed wrongly since it is too close to a heater" (action $a_1$), then he will win with non-absolute strength:

$$M!^{1,4}_{aud}(p_{t<10}) \rightarrow \Box(prop : a_1)(M!^{3,4}_{aud}(p_{t<10})).$$

We read this formula as follows "If the audience believes the claim with the degree $\frac{1}{4}$ then always after the execution of the action $a$ by the proponent, the audience will believe the claim with the degree $\frac{3}{4}$". On the other hand, if the proponent will perform a nonverbal persuasive action moving the thermometer to another place (action $a_1$) (proving this way that the temperature is lower than $10^0$C), he may obtain audience's utter conviction:

$$M!^{1,4}_{aud}(p_{t<10}) \rightarrow \Diamond(prop : a_2)(M!^{4,4}_{aud}(p_{t<10})).$$

The next property of persuasion is the **length and order of argument sequence**. Say that the proponent is able to convince the audience to believe the claim in the degree $\frac{3}{4}$ with support of argument sequence $a_1; a_2; a_3$. The questions we may want to ask to the Perseus system are, firstly, whether it is possible to convince the audience using fewer than three arguments and obtain exactly the same result (or possibly even better), and, secondly, whether it is possible to obtain a better result performing these arguments in a different order, e.g. $a_3; a_1; a_2$? The first question is expressed by such a formula:

$$\Diamond(prop : a_1; a_2; a_3)(M!^{3,4}_{aud}(p_{t<10})) \wedge \Diamond(prop : a_4)(M!^{3,4}_{aud}(p_{t<10})).$$

This formula means that "If the proponent gives three arguments $a_1, a_2, a_3$ then the audience will believe the claim with the degree $\frac{3}{4}$ and if the proponent gives only one argument $a_4$ then the audience will believe the claim with the same degree of $\frac{3}{4}$". The second question is expressed by such a formula:

$$\Diamond(prop : a_1; a_2; a_3)(M!^{3,4}_{aud}(p_{t<10})) \wedge \Diamond(prop : a_3; a_1; a_2)(M!^{4,4}_{aud}(p_{t<10})).$$

This formula says that "If the proponent gives three arguments $a_1, a_2, a_3$ then the audience will believe the claim with the degree $\frac{3}{4}$ and if the proponent order them differently, i.e. $a_3; a_1; a_2$, then the audience will believe the claim with the higher degree of $\frac{4}{4}$".

Persuasiveness can be also affected by the **credibility of a proponent**. Assume that the audience finds a proponent $prop_1$ unreliable. As a result, it does not trust what the proponent says or acts and therefore none of $prop_1$'s arguments will convince *aud*. On the other hand, if another proponent $prop_2$ is a leader of a group of agents or a specialist, then his arguments may have great persuasive power. In other words, the same arguments can cause different results depending on an agent who performs them:

$$\neg\Diamond(prop_1 : P)(M!^{4,4}_{aud}(p_{t<10})) \wedge \Diamond(prop_2 : P)(M!^{4,4}_{aud}(p_{t<10})).$$

### 5.3. Verbal and non-verbal persuasive actions

In [5], the syntax and semantics of $\mathcal{AG}_n$ logic is enriched to allow the representation and verification of properties related to the type of actions performed in persuasion. Every persuasive action is described by 3-tuple $(m, \beta, \delta)$ which fixes a content of a message $m$ sent in the action, a goal $\alpha$ of executing action and the way it is performed. Formally, the set of persuasive actions $\Pi_p \subseteq \Pi_0$ is defined as follows:

$$\Pi_p = \{(m, \beta, \delta) : m \in C, \beta \in F, \delta \in \Delta\}$$

where $C$ is a set of contents, $F$ is a set of formulas of $\mathcal{AG}_n$ and $\Delta$ is a set of symbols representing means of actions, i.e., ways they can be performed. This set can consist of the elements such as: *ver* – for verbal actions, *nver* – for nonverbal actions.

Consider the example given in [8]. John prepares shrimps for a dinner. Mary wants him to add some curry and says "Don't you think these shrimps need curry?". John is not convinced that it will make shrimps taste better and refuses to do it. Mary quits trying to persuade him verbally and tries nonverbal strategy. She goes to the kitchen cupboard, climbs onto the step-stool and begins searching through the upper shelves of the cupboard. Finally, she goes down with a smile and gives him a can of powder. John looks at her and says "Well, yeah, sure, O.K...". The effort that Mary put in finding curry finally made him add curry. In other words, the physical actions resulted in success while verbal action failed. Observe that both arguments – verbal and nonverbal – sends (more or less) the same message $m$ of how important Mary thinks the curry is. That is, the **means of sending** the message can give different results in persuading a receiver of that message.

We say that a goal of an action is achieved if after execution of this action a state, in which it is satisfied, is reached (a formula expressing the goal is true in this state). For example, suppose an action $a = (m, \beta, \delta)$ executed by Mary. The goal of this action is achieved if there exists a state $s$ such that $s$ is a result of $a$ and $s \models \beta$.

Perseus can verify if the same content sent with the same goal but by different means (verbal vs. non-verbal) brings about different results:

$$\Diamond((m, M!^{1,1}_{John}(p), \boldsymbol{nver}) : Mary) M!^{1,1}_{John}(p) \wedge$$
$$\neg\Diamond((m, M!^{1,1}_{John}(p), \boldsymbol{ver}) : Mary) M!^{1,1}_{John}(p).$$

where $p$ means that "shrimps with curry are better". The property expresses that Mary's nonverbal action of sending $m$ is successful, while verbal action of sending $m$ is not. Moreover, it is possible to test which content can cause the success assuming that the goal and means are the same:

$$\Diamond((\boldsymbol{m}_1, \beta, \delta) : i)\beta \wedge \neg\Diamond((\boldsymbol{m}_2, \beta, \delta) : i)\beta.$$

### 5.4. Playing a dialogue game

In [3], we extend the $\mathcal{AG}_n$ logic to allow the formulation of questions referring to the properties of persuasion in agent dialogue games. Consider a dialogue given in [12]:

Paul: My car is safe. (making a claim)

Olga: Why is your car safe? (asking grounds for a claim)

Paul: Since it has an airbag. (offering grounds for a claim)

Olga: That is true, (conceding a claim) but this does not make your car safe. (stating a counterclaim)

Paul: Why does that not make my car safe? (asking grounds for a claim)

Olga: Since the newspapers recently reported on airbags expanding without cause. (stating a counterargument by providing grounds for the counterclaim)

Paul: Yes, that is what the newspapers say. (conceding a claim) OK, I was wrong that my car is safe. (retracting a claim)

The dialogue actions will move the system to a new model in which the sets of commitments will be enriched with new formulas. In our approach this dialogue is a sequence of the following actions:

$((Paul : claim(p)); (Olga : why(p)); (Paul : argue(q, q \rightarrow p; p));$
$(Olga : concede(q)); (Olga : claim(\neg p)); (Paul : why(\neg p));$
$(Olga : argue(r, r \rightarrow \neg p; \neg p)); (Paul : concede(r)); (Paul : retract(p))).$

Perseus will be able to verify that the dialogue satisfies a property. For example it can check whether the formula

$$\Diamond(Paul : claim(p))(\mathbf{C}_{Paul}p \wedge \Diamond(d)\neg\mathbf{C}_{Paul}p)$$

is true at the state $s$ of the model M. $\Diamond(d)$ is the abbreviation for:

$\Diamond(Olga : why(p))\ \Diamond(Paul : argue(q, q \rightarrow p; p))\Diamond(Olga : concede(q))$
$\Diamond(Olga : claim(\neg p))\ \Diamond(Paul : why(\neg p))(Olga : argue(r, r \rightarrow \neg p; \neg p))$
$\Diamond(Paul : concede(r))\ \Diamond(Paul : retract(p)).$

The formula says that at the beginning Paul announces that his car is safe and after the dialogue $d$ he withdraws from this statement.

The Perseus tool may also find the dialogue after which Paul is not committed to the proposition $p$. It is done by parametric verification of the expression

$$\Diamond(Paul : claim(p))(\mathbf{C}_{Paul}p \wedge \Diamond(?)\neg\mathbf{C}_{Paul}p).$$

In other words, Perseus will look for a sequence of actions such that if it is replaced with the symbol ? then this expression becomes a formula true at state $s$ of M [6]. The parametric verification requires searching the whole dialogue game, i.e., all possible dialogues allowed by a given protocol.

### 5.5. Strategies and victory in a persuasion dialogue

In [4], the $\mathcal{AG}_n$ logic is extended to enable the formal verification of strategies allowing an agent to become victorious in a dialogue game. A victory can be specified in different ways, e.g. it may be assumed that the proponent $i$ is the winner if the opponent has conceded $i$'s main claim and the opponent is the winner if the proponent $i$ has retracted $i$'s main claim [12]. Let $win(i)$ mean that $i$ is a winner of a given dialogue game, $t$ be a topic (i.e. a conflict formula), *prop* – a proponent and *opp* – an opponent. Then: (1) $win(prop)$ is true in a state, in which $\mathbf{C}_{opp}(t)$ holds, and (2) $win(opp)$ is true in a state, in which $\neg\mathbf{C}_{prop}(t)$ holds. Using a given specification for a victory, we can ask: is it possible that after performing a dialogue $d$ between $i$ and $\bar{i}$ played in accordance with the protocol $P$, it will be the case that the proposition $win(i)$ will hold.

A more interesting question would be to ask **which dialogue** (sequence of moves) allows an agent to win a dialogue game. This question requires our model checker to perform parametric verification by searching for a legal dialogue (a dialogue played according to a given protocol) such that it is possible that after performing it the proposition $win(i)$ will hold. The answer to this question would allow an agent to plan how to play the dialogue game, however, it has a limitation. Unlike some other types of sequences of actions, a dialogue always consists of actions executed not only by one agent, but also by his adversary. It means that part of a sequence is not under control of a given agent. As a result, even though $i$ knows that a particular sequence leads him to the victory, this sequence may not be performed in a dialogue, since $\bar{i}$ may execute the action allowed by a dialogue protocol, but other than considered by $i$. Say that a sequence *claim p*; *why p*; *p since q*; ...; *concede s* allows $i$ to win a dialogue. Yet, in the second move $\bar{i}$ may execute *claim* $\neg p$ instead of *why p*.

The important property of a persuasion dialogue game is an existence of a strategy which allows an agent to win the game. A strategy for an agent $i$ can be defined as a function from the set of all finite legal dialogues in which $i$ is to move into $L_c$ [13]. Intuitively, $i$ has a strategy if he has a plan of how to react to any move of his adversary. Say that the first move *claim p* is performed by $i$. At this state, $i$ considers how he will respond after all possible moves that $\bar{i}$ is allowed to make at the next stages of a dialogue. In particular, he may plan that at the subsequent stage if $\bar{i}$ executes *why p*, then his response will be: *p since q* (instead of, e.g., *retract p*), if $\bar{i}$ executes *claim* $\neg p$, then his response will be: *why* $\neg p$ (instead of, e.g., *concede* $\neg p$), and so on. An agent may want to know if a strategy that he adopted guarantees him victory in a given dialogue game regardless of what actions his opponent will perform, i.e. if his strategy is winning. A strategy is a **winning strategy** for $i$ if in every dialogue played according to this strategy $i$ accomplishes his dialogue goal [13].

The question about a winning strategy has some limitations, since in some systems of persuasion dialogue a player may avoid losing simply by never giving in [12, p. 1021], e.g. an opponent may repeat *why* $\alpha$ as a response to any assertion that the proponent performs, such as *claim* $\alpha$ or $\beta$ *since* $\alpha$. In such cases, we may want to ask if a strategy allows an agent to reason not about the guarantee but about the possibility of victory. Intuitively, a **strategy gives $i$ chance for success** if there is a dialogue game played according to this strategy such that $i$ accomplishes his dialogue goal. Knowing that the strategy has this feature allows an agent to make decision about which strategy he should adopt in order to have a chance to win. Even though the agent is not sure if he will win, the information that one strategy can bring him success and the other cannot is better than no information.

This type of question has also some limitations. Say that an agent knows that ten strategies allow him to be victorious. How can he decide which strategy to choose? Thus, an agent may wish to know in how many cases a strategy gives him a chance to win. Assume a class of dialogue games in which there is a finite number of possible game's scenarios. Let $k_2$ be the number of all dialogues played according to a given strategy, and $k_1$ be a number of dialogues played according to this strategy in which a given agent $i$ accomplishes his dialogue goal. If $k_1$ and $k_2$ are finite, then we say that this **strategy gives $i$ chance for success in a degree $\frac{k_1}{k_2}$**. If they are infinite, the degree of chance for success is not defined. Knowing that ten strategies allows him to be victorious and knowing their degrees of chance for success allows an agent to choose among them and, in consequence, to maximize his chance to win.

Formally, the properties of dialogue systems concerning the strategies in a dialogue game can be specified and verified with the use of the strategy operators, e.g.

– $\langle\langle i \rangle\rangle$ *true* U *win*($i$) – there is a strategy for agent $i$ which ensure that $i$ will win the dialogue game, i.e. there exists a winning strategy in a game,

- $\langle i \rangle^{0.5}$ *true* U *win*($i$) – there is a strategy which allows an agent $i$ to win a dialogue game with degree higher than 0.5,
- $\langle\langle i \rangle\rangle$ *true* U $(M!^{1,1}_{\bar{i}})t$ – there is a strategy for agent $i$ which ensure that $i$'s adversary will believe the topic $t$ with degree $\frac{1}{1}$,
- $\langle\langle i \rangle\rangle$ *true*U $C_{\bar{i}}(t)$ – there is a strategy for $i$ which ensure that $i$'s adversary will be committed to the topic $t$,
- $\langle\langle prop \rangle\rangle$G(*terminate* $\rightarrow$ *win*(*prop*)) – the proponent has such a strategy that he is a winner every time when the dialogue terminates (i.e. the proposition *terminate* is true),
- $\langle prop \rangle$G(*terminate* $\rightarrow$ *win*(*prop*)) – the proponent has such a strategy that if the dialogue terminates then he may be a winner.

## 6. Conclusions

The paper presents the Perseus model checker which allows the formal verification of persuasion in multi-agent systems. Perseus is built upon the $\mathcal{AG}_n$ logic and performs both the standard model checking method and the parametric verification. In the first case, the tool checks if a given $\mathcal{AG}_n$ formula is true in a given model, while in latter case it searches for answer to a question about a given property of persuasion in a multi-agent system. Formally it means that for an $\mathcal{AG}_n$ expression with unknowns Perseus searches for such values that if the unknowns are replaced with those values, the expression becomes true in a given state and a given model. Perseus is designed and adjusted to verify different properties of persuasion related to influence of persuasion on agent uncertainty, conflict of opinion which initiates persuasion, persuasiveness, the type of means of sending a persuasive message (verbal or non-verbal), playing a persuasion dialogue game and strategies allowing an agent to wine a dialogue game.

## Acknowledgments

Katarzyna Budzyńska gratefully acknowledges the support from Polish Ministry of Science and Higher Education under grant N N101 009338.

## References

[1] E. Black and A. Hunter. An inquiry dialogue system. *Autonomous Agents and Multi-Agent Systems*, 19(2): 173–209, 2009.

[2] K. Budzyńska and M. Kacprzak. A logic for reasoning about persuasion. *Fundamenta Informaticae*, 85: 51–65, 2008.

[3] K. Budzyńska and M. Kacprzak. Formal framework for analysis of agent persuasion dialogue games. In *Proc. of Concurrency, Specification and Programming*, pages 85–96, 2010.

[4] K. Budzyńska and M. Kacprzak. A logic for strategies in persuasion dialogue games. *to be published*, 2011.

[5] K. Budzyńska, M. Kacprzak, and P. Rembelski. Logic for reasoning about components of persuasive actions. *Proc. of ISMIS'09 (18th International Symposium on Methodologies for Intelligent Systems)*, LNAI (5722): 201–210, 2009.

[6] K. Budzyńska, M. Kacprzak, and P. Rembelski. Perseus. software for analyzing persuasion process. *Fundamenta Informaticae*, 2009.

[7] K. Budzyńska, M. Kacprzak, and P. Rembelski. Update of probabilistic beliefs: implementation and parametric verification. *Fundamenta Informaticae*, 102(1): 35–48, 2010.

[8] M. A. Gilbert. Modes, coalescence and pragma-dialectics. http://www.yorku.ca/gilbert/argthry/argthry/arg_readings.html, 1995.

[9] M. Kacprzak, W. Nabiałek, A. Niewiadomski, W. Penczek, A. Półrola, M. Szreter, B. Woźna, and A. Zbrzezny. Verics 2008 – a model checker for time petri nets and high-level languages. In *Proc. of the International Workshop on Petri Nets and Software Engineering (PNSE'09)*, pages 119–132. University of Hamburg, Department of Informatics. Paris, France, 2009.

[10] K. Larson and I. Rahwan. Welfare properties of argumentation-based semantics. In *Proceedings of the 2nd COMSOC*, 2008.

[11] A. Lomuscio, H. Qu, and F. Raimondi. MCMAS: A model checker for the verification of multi-agent systems. In *Proc. of the 21th International Conference on Computer Aided Verification (CAV 2009)*, volume 5643 of *Lecture Notes in Computer Science*, pages 682–688. Springer, 2009.

[12] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, (15): 1009–1040, 2005.

[13] H. Prakken. Formal systems for persuasion dialogue. *The Knowledge Engineering Review*, 21: 163–188, 2006.

[14] I. Rahwan, K. Larson, and F. Tohme. A characterisation of strategy-proofness for grounded argumentation semantics. In *Proceedings of the 21st IJCAI*, 2009.

[15] C. D. Walton. Verifiable agent dialogues. *Journal of Applied Logic*, 5: 197–213, 2007.

[16] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of N.Y. Press, 1995.

Katarzyna Budzyńska
Institute of Philosophy
Cardinal Stefan Wyszyński University in Warsaw
ul. Dewajtis 5, 01-815 Warsaw, Poland
k.budzynska@uksw.edu.pl

Magdalena Kacprzak
Faculty of Computer Science
Biaystok University of Technology
ul. Wiejska 45A, 15-351 Białystok, Poland
mkacprzak@ii.pb.bialystok.pl

http://perseus.ovh.org/
http://argumentacja.pdg.pl/

**Douglas Walton**
University of Windsor

# HOW TO REFUTE AN ARGUMENT USING ARTIFICIAL INTELLIGENCE

**Abstract**: There is a family of terms in argumentation that are closely related to each other and that all refer to some way in which a given argument is attacked, rebutted, refuted, undercut, critically questioned or objected to, thereby defeating it or casting it into doubt. Proper understanding of this family of terms is fundamental to argumentation theory and to building argumentation technologies in artificial intelligence. This paper refines, clarifies and classifies them, using the Carneades Argumentation System. It begins with a simple example that illustrates two main ways of refuting an argument, and concludes with a seven-step procedure for seeking a refutation or objection.

**Keywords**: attack, rebuttal, refutation, challenge, defeater, undercutting defeater, rebutting defeater, exception, objection, Carneades Argumentation System.

This paper applies a computational model to examples of argument attack, challenge, critical questioning and rebuttal, in order to study techniques for refuting an argument. The aim is to improve our understanding of how to attack and refute an argument by clarifying a group of related terms including 'attack', 'rebuttal', 'refutation', 'challenge', 'defeater', 'undercutting defeater', 'rebutting defeater', 'exception' and 'objection' that are commonly used in the literature on argumentation and artificial intelligence. One special kind of objection that is studied is that of irrelevance. As shown in the paper, these terms are, at their present state of usage, not precise or consistent enough for us to helpfully differentiate their meanings in framing useful advice on how to attack and refute arguments. To help remedy this situation, a classification system comprising all these key terms is built and defended.

The term 'rebuttal' is often associated with the work of Toulmin (1958), while the terms 'undercutting defeater' and 'rebutting defeater' are associated with the work of Pollock (1995), and are commonly used in the AI literature. For this reason, section 1 and section 7 of this paper are given over to terminological discussions of these key terms. In section 2, however, we get down to the main job by presenting and working with an example

ISBN 978–83–7431–305–6 ISSN 0860-150X 

meant to illustrate two fundamentally different ways of refuting an argument. One of the attack procedures is called an internal refutation and the other is called an external refutation. It is this distinction, and the example of its use, that provide the departure point for the rest of the paper. In section 3 argumentation schemes with matching sets of critical questions are introduced, using the example of the scheme for argument from expert opinion. In section 4 a computational system from artificial intelligence called Carneades is introduced, and in section 5 and 6 it is shown how Carneades models rebuttal and refutation. Carneades also has a working graphical user interface that is used to visualize arguments and refutations, and this tool is used to analyze the arguments, rebuttals and refutations that are studied. A summary of the seven-step practical procedure for attacking and refuting an argument is given at the end.

## 1. Questions about Attack, Rebuttal, Objection and Refutation

One finds it to be a widely held commonplace in writings on logic and artificial intelligence that there are three ways to attack an argument (Prakken, 2010, 169). One is to argue that a premise is false or insufficiently supported. Let's call this premise attack. Another is to argue that the conclusion doesn't follow from the set of premises that were presented as supporting it. This could be called an undercutting attack, as we will see below. The third is to argue that the conclusion is shown to be false by bringing forward a counter-argument opposed to the original argument. What the attacker needs to do in such a case is to put forward a second argument that is stronger than the original argument and that provides evidence for rejecting the conclusion of the original argument. Such an attack is sufficient to defeat the original argument, unless its proponent can give further reasons to support it.

The undercutting type of attack does not apply to deductively valid arguments. If an argument fits the form of a deductively valid argument, it is impossible for the premises to be true and the conclusion false. Deductive reasoning is monotonic, meaning that a deductive argument always remains valid even if new premises are added. However there is a method of attack on defeasible arguments that is highly familiar in the recent research on nonmonotonic logics for defeasible reasoning. It is to argue that there is an exception to the rule, and that the given case falls under the category of this type of exception. This way of attacking an argument is very familiar in recent studies of defeasible reasoning, like the classical Tweety inference:

birds fly; Tweety is a bird; therefore Tweety flies. This inference is based on the defeasible generalization that birds normally fly, or it could also be analyzed as being based a conditional rule to the effect that if something is a bird it flies. Such a conditional is open to exceptions, meaning that it may default in some cases. The argument can be attacked by pointing out the exception to the rule.

To attack an argument in the third way, it may be enough to simply question whether its conclusion is true, but if a given argument that is being attacked has a certain degree of strength, merely questioning its conclusion may not be sufficient. What the attacker needs to do in such a case is to put forward a second argument that is stronger than the original argument and that provides evidence for rejecting the conclusion of the original argument. Such an attack is sufficient to defeat the original argument, unless its proponent can give further reasons to support it (Dung, 1995). Still another way to attack an argument is to ask a critical question that casts the argument into doubt, and that may defeat the argument unless its proponent can make some suitable reply to the question. The form of attack will be taken up in section 4.

Even though the given argument may stand, having repelled all attacks of the first three kinds, it may still be defeated on other grounds. One of these is that the argument is irrelevant, even though it may be valid. What is presupposed by this fourth kind of attack is that the given argument is supposed to be used to resolve some unsettled issue in a discussion that is being carried on in the given case. To attack an argument in the fourth way, matters of how the argument was used for some purpose in a context of dialogue need to be taken into account. If an argument has no probative value as evidence to prove or disprove the ultimate *probandum* in this particular discussion, in may be dismissed as irrelevant. Discussions of argument attack and refutation in the literature tend to acknowledge the first three ways of attacking an argument but to overlook the fourth way. The reason could be that this fourth way is more contextual than the first three ways in that it more directly relates to the context of dialogue surrounding the given argument. It could be classified as a procedural objection rather than as an attack.

Still another way to attack an argument is to claim that it commits the fallacy of begging the question. A circular argument, like 'Snow is white therefore snow is white', may be deductively valid but still be open to attack on the grounds that it fails to prove its conclusion. The failure here relates to the requirement that the premises of an argument that is being used to prove a conclusion should carry more weight than the conclusion itself. Thus if one of the premises depends on the conclusion, and cannot be proved independently of the conclusion, it is useless to increase the probative weight of the conclusion. Such an argument may be valid, but it is open to the criticism that it is useless to prove the conclusion it is supposed to be proving.

Although there may be four basic ways to attack an argument, asking a critical question is a way of making an objection to an argument that may, or may not be seen as an attack on the argument. The notion of making an objection to an argument seems to be much broader than the notion of attacking an argument, for making an objection can be procedural in nature. We also need to be careful to note that there can be ways of making an objection to an argument that do not fall into any of these five categories of attack on an argument (Krabbe 2007). Thus the task of defining the notion of an objection precisely, and the task of classifying the various types of objections that can be made to an argument, remain open questions for future work. Still, in this section we have made some progress towards this investigation by carefully describing four basic ways to attack an argument, and by adding that asking a critical question may also often be seen as a way of attacking an argument by raising critical doubts about it. Argument attacks surely represent some of the central ways of raising an objection about an argument.

Perhaps the best known use of the term 'rebuttal' in argumentation theory is Toulmin's use of it in his argument model, containing the elements datum, qualifier, claim, warrant, backing and rebuttal. In the model (Toulmin, 1958, 101), the datum is supported by a warrant that leads to a claim that is qualified by conditions of exception or rebuttal. For example (99), the claim that a man is a British subject might be supported by the datum that he was born in Bermuda, based on the warrant that a man born in Bermuda will be a British subject. The warrant appears to be similar to what is often called a generalization in logic. This example of an argument is defeasible, because the generalization is subject to exceptions, and hence the argument is subject to defeat if the information comes in showing that the particular case at issue is one where an exception holds. For example, although a man may have been born in Bermuda, he may have changed his nationality since birth (101). Toulmin uses the word 'rebuttal', but other words like 'refutation' or 'defeater' might also be used to apply to such a case.

The meaning term of the term 'warrant' in Toulmin's argument layout has long been the subject of much controversy (Hitchcock and Verheij, 2006). A Toulmin warrant is in typical instances a general statement that acts as an inference license, in contrast to the datum and claim that tend to be specific statements. In logical terms, it could be described as a propositional function or open sentence of this form: if a person $x$ was born in Bermuda, then generally that person $x$ is a British subject.

A rebuttal, judging by Toulmin's Bermuda example, is an exception to a rule (warrant, in Toulmin's terms). However, according to Verheij (2008, 20), rebuttal is an ambiguous concept in Toulmin's treatment, and five meanings of the term need to be distinguished. First, rebuttals are associated with "circumstances in which the general authority of the warrant would have to be set aside" (Toulmin, 1958, 101). Second, rebuttals are "exceptional circumstances which might be capable of defeating or rebutting the warranted conclusion" (Toulmin, 1958, 101). Third, rebuttals are associated with the non-applicability of a warrant (Toulmin, 1958, 102). But a warrant could also be an argument against the datum, a different sort of rebuttal from an argument against the warrant or the claim. In traditional logical terms, this would be an argument claiming that a premise of the inference being rebutted does not hold. Verheij also distinguishes between the warrant that acts as an evidential support of the conditional and the conditional that is one premise in the inference. On his analysis a rebuttal can attack the conditional or it can attack the warrant that supports the conditional as evidence.

Describing rebuttal as citing an exception to a rule of inference on which an argument was based sounds similar to what is called undercutting in the literature on defeasibility (Pollock, 1995). Pollock's distinction between two kinds of counter-arguments called rebutting defeaters and undercutting defeaters (often referred to as rebutters versus undercutters) is drawn as follows. A rebutting defeater gives a reason for denying a claim by arguing that the claim is a false previously held belief (Pollock, 1995, p. 40). An undercutting defeater attacks the inferential link between the claim and the reason supporting it by weakening or removing the reason that supported the claim. The way Pollock uses these terms, a rebutter gives a reason to show the conclusion is false, whereas an undercutter merely raises doubt whether the inference supporting the conclusion holds. It does not show that the conclusion is false. The classic example is the Tweety argument: Birds fly, Tweety is a bird; therefore Tweety flies. If new information comes in telling us that Tweety is a penguin, the original Tweety argument is undercut. Generally speaking, the argument still holds. Generally birds fly, and hence, given that Tweety is a bird, it follows that Tweety flies. But in this particular case, we have found out that Tweety is a penguin. Hence in this particular case, since we know that Tweety is type of bird that does not fly, we can no longer use the former inference to draw the conclusion that Tweety flies.

Pollock has another example (1995, 41) that illustrates a defeasible argument that could be called argument from perception.

> For instance, suppose $x$ looks red to me, but I know that $x$ is illuminated by red lights and red lights can make objects look red when they are not. Knowing this defeats the prima facie reason, but it is not a reason for thinking that $x$ is *not* red. After all, red objects look red in red light too. This is an *undercutting defeater* (Pollock's italics in both instances).

To show how the red light example has the defining characteristics of a species of rebuttal, we can analyze it as an initial (given) argument and a counter-argument posed against it. The original argument says: when an object looks red, then (normally, but subject to exceptions) it is red, and this object looks red to me, therefore this object is red. The rebuttal of the original acts as a counter-argument that attacks the original argument: this object is illuminated by a red light, and when an object is illuminated by a red light, this can make it look red even though it is not, therefore the original argument (the *prima facie* reason for concluding that this object is red expressed by the original argument) no longer holds. According to Pollock (1995, 41) the counter-argument should be classified as an undercutter rather than a rebutter because red objects look red in red light too. Even given the attacking argument, the object may be red, for all we know. Thus in Pollock's terms it would not be right to say that the attacking argument is a rebutting defeater that shows that the conclusion of the original argument is false. What it shows is that because of the new information about the red light, the counter-argument, built on this new information, casts doubt on the conclusion of the original argument. As an undercutter it acts like a critical question that casts an argument into doubt.

Pollock's distinction between rebutters and undercutters is clearly a fundamental to any understanding of defeasible reasoning, but from a practical point of view, it leaves a number of questions open. Is an undercutter a particular instance that makes a defeasible generalization fail in a specific case? Or is an undercutter a special type of counterargument that attacks a prior defeasible argument and acts as a rebuttal to it? Is there a special characteristic of the logical structure of defeasible arguments that leaves them open to an undercutter type of attack, and if so how can we identify this characteristic so that we can learn when making an undercutter type of attack is appropriate? These are all practical questions that seek guidance that might be helpful in telling a participant in argumentation, or a critic of an argument, how to attack that argument or critically question it by finding some sort of standard rebuttal that applies to it. There are also some terminological questions about how to classify the terms ‘attack’, ‘rebuttal’ and ‘refutation’. Pollock’s terminology can be somewhat confusing when we try to apply it to giving practical advice on how to attack, rebut, critically question or refute a given argument, because undercutting does not sound all that different from rebutting. If I find an exception to a rule that defeats the defeasible argument, as in the red light example, surely it is reasonable to say that I have rebutted the original argument. We could even take a step further and use a stronger word, saying that I have refuted the argument. How is rebuttal different from refutation, a term often used in logic textbooks and writings on logic over the centuries? To approach these questions it is best to begin with a practical, and apparently simple example, in which advice is given on how to refute an argument, or perhaps as we might say, to rebut or attack an argument.

## 2. Internal and External Refutation

Goodwin (2010) presented a methodical procedure to her students on how to refute an argument that contrasts two strategies. The first strategy is that of focusing on the argument’s conclusion and arguing for the opposite. She offered the following example. If one side argues that video games lead to violence, the other side can argue that video games do not lead to violence. This can be recognized as a strategy often called rebuttal or refutation. It is the strategy when confronted with a target argument to present a new argument that has the opposite (negation) of the target argument as its conclusion. Although conceding that this is an important and often effective strategy, she suggests another one that may be even better. Instead of just looking at the conclusion of the other argument, this second strategy is to examine the reasons the other side is giving to support its argument, and see if these reasons hold up under questioning. Among the questions she proposed as ways of attacking the other argument are (1) to ask whether the other side is relying on a biased source, (2) to ask whether the evidence the other side is citing is relevant, or (3) to ask whether the analogy put forward by the other side is really similar.

What is suggested by this advice is that there are basically two ways of attacking an argument. One way, generally called refutation, is to present a new argument that has as its conclusion the negation of the original argument.[1] The other way, generally called asking critical questions, or casting doubt on an argument, is to ask questions that relate to the particular form of the original argument. For example, if the original argument was based on a source, like witness testimony or expert testimony, one could ask the critical question of whether that source is biased. Or if the original argument has the form of an argument from analogy, one could ask the critical question of whether the two cases at issue are really similar. Goodwin states that although attacking the other side's reasons by asking critical questions involves more strategy and paying attention to what the other side says, it can often be more effective because it attacks the opposed argument internally, nicely causing it to fall down.

This practical advice on how to refute an argument is generally very interesting from the point of view of argumentation theory, because it suggests there are two distinctive strategies, refutation and critical questioning, as each might be called, that need to be separated, and that each calls for a different approach. She has shown that each type of argument strategy has a distinctively different structure from the other. This is an important distinction for argumentation theory. Hamblin (1970, 162) distinguished between a weaker and a stronger sense of the term 'refutation'. The weaker he describes as "destruction of an opponent's proof" and the stronger as "construction of the proof of a contrary thesis". It would be nice to have some terminology to make this important distinction between these two meanings of the term 'refutation'. Let us call destruction of an opponent's proof internal refutation, because, as Goodwin has described it, this strategy is to examine the reasons the other side is giving to support its argument, and see if these reasons hold up under questioning. It is an internal attack on the argumentation offered by the other side. Let us call the construction of the proof of a contrary thesis external refutation because it goes outside the original argument to present a new argument that has as its conclusion the negation of the original argument. Attacks can be internal or external.

An example she gives to illustrate the technique of internal refutation is quoted below (with some parts deleted).

[1] Below we will challenge this generally accepted meaning of the term 'refutation' on the grounds that it is too broad. The problem is that we often have cases where a new argument has as its conclusion the negation of an original argument, but the new argument might still be weaker than the original argument. In such case it is questionable whether the new argument is a refutation of the original one. For the moment, however, we accept the broad conventional meaning of the term 'refutation' as a point of departure.

> The other side said that Dr. Smith's study clearly shows that video games do not lead to violence. But Dr. Smith is biased. His research is entirely funded by the video game industry. That's what the 2001 investigation by the Parent's Defense League demonstrates. So you can see that the other side has no credible evidence linking video games to violence.

In the example one can see the components of a refutation. First, there are two parties that are presenting arguments on opposed sides of the disputed issue. The issue is whether or not video games lead to violence. The first side has argued that video games do not lead to violence, and has supported its claim by bringing forward the evidence that Dr. Smith's study shows that this claim is true. The opposed side then presents a counterargument, but this counterargument is not an external refutation, a new argument that supports the claim that video games do lead to violence. Instead, it attacks the original argument internally by making the claim that Dr. Smith is biased, and supports it with the reason that his research is entirely funded by the video game industry. So this is a counterargument, but not a refutation in the sense defined above. It is something else. It corresponds to the other technique of attacking an argument that Goodwin described as attacking the reasons the other side is giving by asking critical questions.

We can even analyze this internal type of attacking strategy more deeply by pointing out that the original argument took a particular form. It appears to be an argument from expert opinion that cites a study by someone called Dr. Smith that supposedly showed that video games do not lead to violence. The field of expertise of Dr. Smith is not stated, but it appears we are meant to assume that Dr. Smith is an expert in some field that includes the study of whether video games lead to violence or not. If we can make this assumption, the form of the original argument can then be identified as that of argument from expert opinion. Given this assumption we can understand a little more about the structure of the internal attack used against this argument. The attack makes the claim that Dr. Smith is biased, and this particular type of attack undercuts the argument by finding a weak point in its structure that, once pointed out and supported by evidence, subjects the argument to doubt in such a way that it no longer holds up as a way of supporting its conclusion that video games do not lead to violence. To understand more about how defeasible arguments can have different forms we need to examine the notion of an argumentation scheme.

### 3. Argumentation Schemes and Critical Questions

Pollock's red light example can be fitted to an argumentation schemes that has been called argument from appearance (Walton, 2006). Although Pollock did not employ the concept of an argumentation scheme with matching critical questions, the pattern of inference of the red light example can be called argument from perception (Walton, Reed and Macagno, 2008, 345).

> PREMISE 1: Person $P$ has a $\varphi$ image (an image of a perceptible property).
>
> PREMISE 2: To have a $\varphi$ image (an image of a perceptible property) is a *prima facie* reason to believe that the circumstances exemplify $\varphi$.
>
> CONCLUSION: It is reasonable to believe that $\varphi$ is the case.

Walton Reed and Macagno (2008, 345) list this form of argument as an argumentation scheme with the following critical question matching it: are the circumstances such that having a $\varphi$ image is not a reliable indicator of $\varphi$?

Consider another example (Prakken, 2003): if something looks like an affidavit, it is an affidavit; this object looks like an affidavit; therefore it is an affidavit. This inference might fail if we are taking part in a TV series about a trial in which props are used. A document on a desk might look like an affidavit, but after all this is a TV series. It might not be an affidavit, but merely a prop made to look like one. In the context, the original argument fails to support the conclusion that the document in question is an affidavit. But maybe it is a real affidavit. An easy way to get such a prop for the TV series would be to ask someone who has access to real affidavits to get one for use in the TV series. This example has the same scheme as the red light example.

The scheme representing argument from expert opinion was formulated in (Walton, 1997, 210), with some minor notational changes, as shown below, with two premises and a conclusion. $E$ is an autonomous agent of a kind that can possess knowledge in some subject domain. The domain of knowledge, or subject domain, is represented by the variable $F$ for field of knowledge. It is assumed that the domain of knowledge contains a set of propositions.

> Major Premise: Source $E$ is an expert in field $F$ containing proposition $A$.
>
> Minor Premise: $E$ asserts that proposition $A$ (in field $F$) is true (false).
>
> ---
>
> Conclusion: $A$ may plausibly be taken to be true (false).

As shown in (Walton, 1997) any given instance of an argument from expert opinion needs to be evaluated in a dialogue where an opponent (respondent) can ask critical questions. This form of inference is defeasible, provided we take it to be based on a defeasible generalization to the effect that if an expert says $A$, and $A$ is in the right field for the expert, then $A$ may plausibly be taken to be acceptable as true (subject to exceptions). What kinds of exceptions need to be taken into account corresponding to critical questions matching a scheme? The six basic critical questions matching the appeal to expert opinion (Walton, 1997, 223) are the following.

$CQ_1$: *Expertise Question*: How knowledgeable is $E$ as an expert source?
$CQ_2$: *Field Question*: Is $E$ an expert in the field $F$ that $A$ is in?
$CQ_3$: *Opinion Question*: What did $E$ assert that implies $A$?
$CQ_4$: *Trustworthiness Question*: Is $E$ personally reliable as a source?
$CQ_5$: *Consistency Question*: Is $A$ consistent with what other experts assert?
$CQ_6$: *Evidence Question*: Is $E$'s assertion based on evidence?

$CQ_1$ refers to the expert's level of mastery of the field $F$. $CQ_4$ refers to the expert's trustworthiness. For example, if the expert has a history of lying, or is known to have something to lose or gain by saying $A$ is true or false, these factors would suggest that the expert may not be personally reliable. The assumption made in (Walton, 1997) was that if the respondent asks one of the six critical questions, the initiative shifts back to the proponent's side to respond to the question appropriately. The asking of the critical question defeats the argument temporarily until the critical question has been answered successfully. This approach was a first pass to solving the problem of how to evaluate an argument from expert opinion. More specifically, it was designed to offer students in courses on critical argumentation some direction on how to react when confronted with an argument from expert opinion. Although the critical questions stated in (Walton, 1997) were meant to be practically useful for this purpose, they are also open to formulation in a more precise manner that might make theoretical refinement possible.

The study of attacks, rebuttals and refutations would be aided considerably if some structure could be brought to bear that would enable us to anticipate in a particular case what sort of attack an argument is susceptible to. Here the critical questions matching a scheme can be very useful. For example if the argument is an appeal to expert opinion, we can see already from examining the critical questions matching scheme for argument from expert opinion that this argument will tend to be open to certain types of attack. For example, it will be open to an attack on the grounds that the expert is not a trustworthy source. One of the standard ways of arguing that an expert is not a trustworthy source is to allege that the expert is biased because she has something financially to gain by making the claim. However, it has been shown that critical questions differ in their force. In some instances, merely asking a critical question makes the original argument default, while in other instances, asking the critical question does not make the argument default unless the question asker can offer evidence to back up the question (Walton and Godden, 2005). There are differences between the critical questions on how strongly or weakly asking the question produces such a shift of initiative. Such observations have led to two theories about requirements for initiative shifting when critical questions matching the argument from expert opinion are asked (Walton and Godden, 2005). According to one theory, in a case where the respondent asks any one of these critical questions, the burden of proof automatically shifts back to the proponent's side to provide an answer, and if she fails to do so, the argument defaults (is defeated). On this theory, only if the proponent does provide an appropriate answer is the plausibility of the original argument from expert opinion restored. According to the other theory, asking a critical question should not be enough to make the original argument default. The question, if questioned, needs to be backed up with some evidence before it can shift any burden that would defeat the argument.

Recent advances in artificial intelligence have developed formal systems to model argumentation, and argument visualization tools that can be used to represent not only reasons given to support an argument, but also attacks on it. Some of these formal systems with visualization tools can also accommodate argumentation schemes. One such system, called Carneades after the Greek skeptical philosopher, can use heuristic strategies to search a space of arguments induced by argumentation schemes (Gordon, 2010). Argumentation schemes in the Carneades model function as heuristic search procedures that apply statements from a database to find arguments pro or con a claim at issue. The arguments that turn up in the resulting stream are alternative ways that can be used to prove the claim. Carneades provides an integrated dialectical framework enabling a variety of legal argumentation schemes, such as argument from expert opinion, to be used in a comprehensive system supporting both argument construction and argument evaluation tasks.

## 4. The Carneades System

Part of the definition of a rebuttal is that it is an attack on an argument, and a rebuttal itself would normally seem to be an argument. In order to define the notion of a rebuttal, we also need to have some clear notion of what an argument is. There is not much agreement in argumentation theory on how to define an argument, however. To cope with this problem, it is best to begin with a minimalist account of the structure of an argument. According to this account, an argument is composed of three things: a set of premises, a conclusion, and an inference that leads from the premises to the conclusion. The conclusion is generally taken to be a claim that has been made, and the premises are propositions that are put forward in support of the claim. Beyond this minimal account, it will prove to be useful to have a formal model to represent the notion of an argument, preferably one that would enable us to visualize the premises and conclusion of an argument in a clear way to represent examples of attacks rebuttals and refutations. For example, if we could represent Goodwin's example of an internal refutation, this capability could be extremely helpful. There a many such argumentation visualization tools available at the present time, but it is especially helpful to use one that provides not only a formal model of argumentation, but also an argument visualization tool that fits the the model.

Carneades is a mathematical model consisting of definitions of mathematical structures and functions on these structures (Gordon, Prakken and Walton, 2007), and a computational model, meaning that all the functions of the model are computable (Gordon and Walton, 2009). Carneades has been implemented using a functional programming language, and has a graphical user interface. (http://github.com/carneades/carneades). Argumentation is modeled by Carneades in a tree structure where the nodes are text boxes containing premises and conclusions of an argument (Gordon, 2010). The premises are connected to the conclusion in the normal way in an argument with an arrow pointing to the conclusion. An argument that supports a conclusion is indicated by a circle containing a + sign. The premise is an exception is joined to a circle by a dashed line. How Carneades displays the structure of the argument in Pollock's red light example is shown in figure 1.

As shown in figure 1, the statement at the bottom right is an exception, and so the argument as a whole represents a Pollock-style undercutter. In the Carneades model, this argument is represented as a typical defeasible argument that has two normal premises, displayed as the top two boxes on the right in figure 1. But this argument is subject to an exception, and in Carneades the exception is represented as an additional premise of a special kind that can defeat the original argument. Carneades can also be used for evaluating arguments, and how the procedure of evaluation can be illustrated using the case of Pollock's red light example is shown in figure 2.

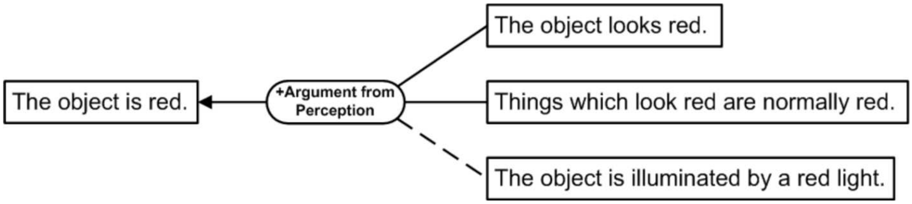



**Figure 1: Exception in Pollock's Red Light Example**

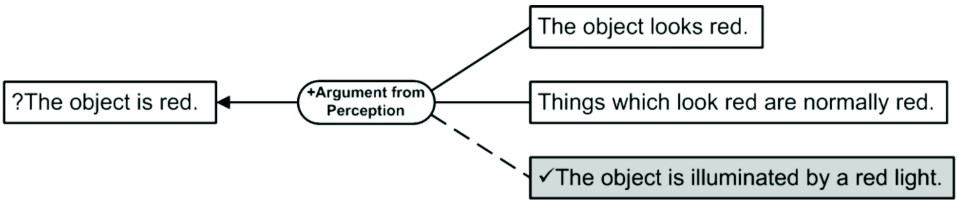



**Figure 2: Undercutter in Pollock's Red Light Example**

As shown by the checkmark text box at the bottom, the statement that the object is illuminated by red light has been accepted. Once the statement has been accepted, even though the two premises above it would normally enable the conclusion to be accepted provided these two premises are accepted, in this situation, since the exception applies, the conclusion is cast into doubt. The status of the conclusion is represented by the question mark appearing in its text box. What this analysis visualizes is a situation in which the conclusion is rendered questionable, and hence not acceptable. It does not tell us, however, that the conclusion is false or unacceptable.

Carneades defines formal properties that are used to identify, analyze, construct, visualize and evaluate arguments (Gordon and Walton, 2006). Part of the definition of a rebuttal is that it is an attack on an argument, and a rebuttal itself is also an argument. It follows that in order to define the notion of a rebuttal, we surely also need to have some clear notion of what an argument is. As noted just above in this section, an argument is taken to have three basic components: a set of premises, a conclusion, and an inference that leads from the premises to the conclusion.

Figures 1 and 2 show how these three components are related. In the following formal definition of an argument in Carneades (Gordon and Walton, 2009), a distinction is drawn between two types of opposition. One is negation, represented in the same way as in classical propositional logic where a proposition $p$ is true if and only if its negation is false. The negation of a proposition, in other words, has the opposite truth value of the original proposition. The other is complement. The complement of a set is the set of things outside that set (Gordon and Walton, 2009, 242–243).

> Let $L$ be a propositional language. An *argument* is a tuple $\langle P, E, c \rangle$ where $P \subset L$ are its *premises*, $E \subset L$ are its *exceptions* and $c \in L$ is its *conclusion*. For simplicity, $c$ and all members of $P$ and $E$ must be literals, i.e. either an atomic proposition or a negated atomic proposition. Let $p$ be a literal. If $p$ is $c$, then the argument is an argument *pro* $p$. If $p$ is the complement of $c$, the argument is an argument *con* $p$.

According to this definition we can understand the notions of an argument pro a proposition $p$ and argument con a proposition $p$ as follows. If $p$ is the conclusion of the argument, the argument is said to be *pro* $p$, whereas if some proposition other than $p$ is the conclusion of the argument, the argument is said to be *con* $p$. Defeaters (rebuttals) are modeled as arguments in the opposite direction for the same conclusion. If one argument is *pro* the conclusion, its rebuttal would be another argument *con* the same conclusion. Premise defeat is modeled by an argument *con* an ordinary premise or an assumption, or *pro* an exception (Gordon, 2005, 56). In the Carneades system, critical questions matching an argument are classified into three categories: ordinary premises, assumptions or exceptions. External refutations are modeled as arguments in the opposite direction for the same conclusion. If one argument is *pro* the conclusion, its refutation would be another argument *con* the same conclusion. Premise defeat is modeled by an argument *con* an ordinary premise or an assumption, or *pro* an exception (Gordon, 2005, 56). See how Carneades models the distinction between internal and external refutation, we show how this distinction works in the case of argument from expert opinion.

Let's begin with the notion of external refutation to see how it works generally in cases of argument from expert opinion. In a case of external refutation, as shown in figure 3, we have one argument from expert opinion in which the premise is that expert 1 says that some proposition $A$ is true and the conclusion is the proposition that $A$ is true. This is the argument shown at the top in figure 3, and it is a pro-argument, as shown by the + in the circle representing the argument. Beneath it is the second argument that attacks the first argument, based on the premise that there is another expert who says that the opposite of $A$ is true. The second argument is an external refutation of the first one, because it is a separate opposed argument that has the opposite conclusion of the first argument. But if the second argument is merely a rebuttal of the first argument can it properly be called a refutation?

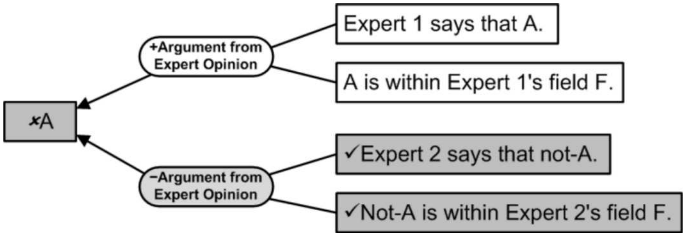



**Figure 3: How Carneades Models External Refutation**

Certainly it fits the definition of an external refutation of the kind attributed to Hamblin above, but there is more to say about it. Notice that in figure 3, the premises shown at the bottom appear in darkened textboxes and have check marks in front of them, indicating that this premise has been accepted. Notice that the premises shown at the top appear in an undarkened text box with no check mark in front, indicating that each premise has merely been stated but has not been accepted. What will happen automatically in Carneades is that the bottom argument will be taken as refuting the top one. Since it has two accepted premises, when both premises are considered together, the conclusion $A$ comes out as rejected (indicated by X).

In such a case, we can say that the first argument is refuted by the second one in a strong sense of the term 'refutation' meaning not only that the second argument goes to the opposite conclusion of the first one, but it does so in such a way that it overwhelms the first argument, providing a reason to infer that the conclusion of the first argument is no longer acceptable. We could say that in this strong sense of refutation, the second argument successfully refutes the first argument. Or perhaps we could draw the distinction in a different way by saying that second argument not only rebuts first argument but also refutes it. The terminology remains uncertain here but we will clarify it later.

No matter how we describe what has happened in this example in terms of the distinction between rebuttal and refutation, we can see why it illustrates how Carneades models the notion of an external refutation. In an external refutation, we have two separate arguments, and one attacks the other externally by providing an independent line of argument that goes to the opposite of the conclusion of the first argument. Carneades models the notion of an internal refutation in a completely different way by focusing on the critical questions matching the argumentation scheme, and goes into considerations of different ways these critical questions can be used to attack the original argument. One of the main features of Carneades is that it enables critical questions to be represented on argument diagrams of the kinds shown in the figures 1, 2 and 3 above (Walton and Gordon, 2005). In the standard argument diagrams, the text boxes (nodes in the tree) contain propositions that are premises and conclusions of arguments, but there is no obvious way that critical questions can be represented on such a diagram. Carneades solves this problem by enabling a distinction to be drawn between two ways an argument from expert opinion should be critically questioned, and thus enables the critical questions to be represented as implicit premises of an argumentation scheme on an argument diagram. The two assumptions that (1) the expert is not trustworthy and (2) that what she says is not consistent with what other experts say, are assumed to be false. It is assumed, in other words, that (1) and (2) are false until new evidence comes in to show that they are true. The two assumptions that (1) the expert is credible as an expert and that (2) what she says is based on evidence, are assumed to be true, until such time as new evidence comes in showing they are false. Also assumed as true are the ordinary premises that (1) the expert really is an expert, (2) she is an expert in the subject domain of the claim, (3) she asserts the claim in question, and (4) the claim is in the subject domain in which she is an expert.

Now let's look once again at the expertise question, to see how it could be classified. It is about $E$'s depth of knowledge in the field $F$ that the proposition at issue lies in. As noted above, the expertise question seems to ask for a comparative rating. What if the proponent fails to answer by specifying some degree of expertise, like "very credible" or "only slightly credible"? As noted above it seems hard to decide what the effect on the original argument should be. Should it be defeated or merely undercut? It seems like it should only be undercut, because even if we don't know how strong the argument from expert opinion is, it might still have some strength. It might even be very strong, for all we know.

The field and opinion questions can be modeled as ordinary premises of the arguments from expert opinion scheme in Carneades. Now let's look back at the trustworthiness question, which refers to the reliability of the expert as a source who can be trusted. If the expert was shown to be biased or a liar, that would presumably be a defeater. It would be an *ad hominem* argument used to attack the original argument, and if strong, would defeat it. But unless there is some evidence of ethical misconduct, as noted above, the proponent could simply answer 'yes', and that would seem to be enough to answer the question appropriately. As noted above, to make such a charge stick, the questioner should be held to supporting the allegation by producing evidence of bias or dishonesty.

According to the discussion above, only the consistency and backup evidence questions need some evidence to back them up before the mere asking of the question defeats the original argument. Hence only these two of the critical questions are treated as exceptions. The results of how the critical questions should be classified as premise on the Carneades model can be summed up as follows.

Premise: $E$ is an expert.
Premise: $E$ asserts that $A$.
Premise: $A$ is within $F$.
Assumption: It is assumed to be true that $E$ is a knowledgeable expert.
Assumption: It is assumed to be true that what $E$ says is based on evidence in field $F$.
Exception: $E$ is not trustworthy.
Exception: What $E$ asserts is not consistent with what other experts in field $F$ say.
Conclusion: $A$ is true.

It is shown in figure 4 how argument from expert opinion is visually represented in the Carneades graphical user interface, and how each premise is represented. A normal premise is represented by a solid line, an exception is represented by a dashed line, and assumption is represented by a dotted line.

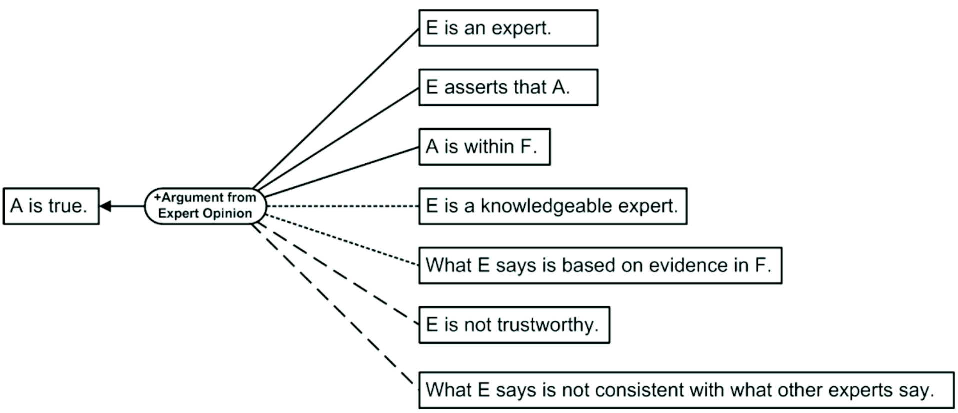



**Figure 4: Visualization of Argument from Expert Opinion in the Carneades Interface**

As figure 4 shows, the critical questions are represented as additional premises alongside the ordinary premises in the scheme for argument from expert opinion. This means that, as far as Carneades is concerned, attacking the argument by asking anyone critical questions can be classified as a premise attack argument. According to the Carneades model, the ordinary premises are stated, whereas the other premises expressing critical questions are either assumptions or exceptions.

If we are using the Carneades graphical user interface to help us devise a strategy to refute an argument we are confronted with, we can look over the evidence available in the case, or that could possibly be collected in the case, in order to decide which of the critical questions would be the best one to pose. Posing a critical question of the assumption type requires no evidence to back it up in order to defeat the original argument. These would be the first premises to look at. Goodwin described the strategy as one of examining the reasons the other side is giving to support its argument to see if these reasons hold up under critical questioning. However, if there is evidence that could be used to back up one of the critical questions, that would be the question to pose. As we see in the case of Dr. Smith, there is evidence that could be used to back up the claim that he is biased. Hence Carneades can automatically point to the trustworthiness question, represented as an exception in the argument visualization, and indicate that the best strategy is to ask this question.

## 5. How Carneades Models Attacks and Rebuttal

Not only are schemes classified under other schemes, but critical questions also have a classification structure as well. For example, although argument from bias is a specific type of argument in its own right with its distinctive argumentation scheme, asking a critical question about bias is so common in responding to arguments from expert opinion that it needs to be identified as a specific critical question in its own right with respect to the scheme for argument from expert opinion. In (Walton 1997, 213–217) the bias critical question is treated as a sub-question of the trustworthiness question. In other words, questioning whether an expert is biased is treated as a special case of questioning whether the expert is personally reliable as a source. The reason is that questioning on grounds of bias is a way of questioning the trustworthiness of an expert source. A biased expert need not be completely untrustworthy, but if there are grounds for suspecting a bias, that is a good reason for having reservations about the strength or even the acceptability of an argument from expert opinion.

Let's go back to the example Goodwin gave to illustrate the technique of attacking the reasons the other side has put forward in its argument. In this example, the attack alleges that Dr. Smith is biased, because his research is entirely funded by the video game industry. Next, evidence to support this claim of bias is put forward. It is claimed that the 2001 investigation by the Parent's Defense League constitutes evidence to support bias. The structure of this argument from expert opinion is shown in figure 5. The three ordinary premises of the argument from expert opinion are shown at the top in the three darkened boxes. In each case, a check mark appears before the proposition in the box. The checkmark redundantly shows, along with the darkened box, that these three propositions have been accepted. Carneades would automatically darken the box for the conclusion and put a checkmark before the proposition that *A* may be taken to be true. This is the normal evaluation procedure that Carneades is set up to automatically carry out. However in this instance, the box in the middle at the bottom containing the proposition that *E* is not trustworthy also has a checkmark in front of it. Moreover, this proposition is supported by evidence of *E*'s bias, and we can take it that this evidence is strong enough to be accepted. Since the premise that *E* is not trustworthy is an exception, the argument from expert opinion to the conclusion that *A* is true is cast into doubt. Hence we see that this proposition has a question mark in front of it in figure 5. Figure 5 shows generally how the trustworthiness and bias critical questions are modeled by Carneades, and how a finding of bias functions as support for a trustworthiness exception of the kind that can cast an argument from expert opinion into doubt by rebutting it.

Next let's examine how Carneades represents the example of the second strategy of refutation described by Goodwin. In this example, it is argued that Dr. Smith's study can be attacked internally by arguing that it was paid for by the video industry. The basic argument is shown in figure 6.

When it comes to evaluating the argument to see how a rebuttal works, we could look at the Carneades visualization of it in figure 6, where the two normal premises at the top are accepted. The remaining premises, as shown in figure 5, are not shown in figure 6. In figure 6 it is also shown that the exception at the bottom, the proposition that Smith was paid by the video industry, is accepted on the grounds that it is supported by the evidence of the 2001 investigation. Although the two premises at the top would normally be enough to support acceptance of the conclusion on the Carneades model, in this instance the conclusion is not accepted. It is

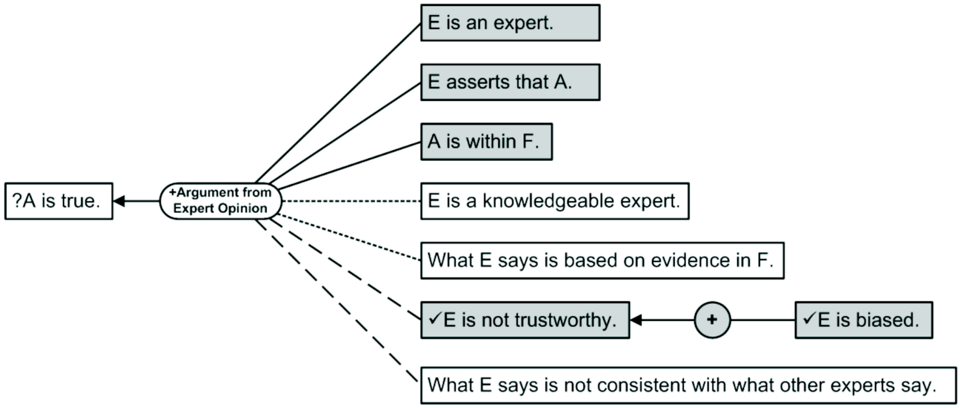



**Figure 5: Argument Undercut by Bias Attack**

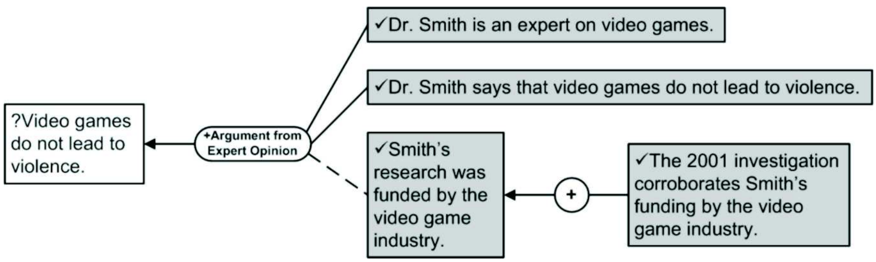



**Figure 6: Defeasible Structure of the Video Example Visualized by Carneades**

shown as questioned. The reason is that the premise at the bottom, the proposition that Smith was paid by the video industry, is an exception, and moreover it is an exception that has been accepted, based on the evidence of the investigation. Thus what figure 6 shows is that the argument from expert opinion has been defeated. It has been undercut by giving evidence to show that an exception applies. Goodwin described the strategy as one of examining the reasons the other side is giving to support its argument to see if these reasons hold up under critical questioning. In this example, it is fair to say that the argument did not hold up under critical questioning. But the question is: has this argument been refuted, or has the conclusion merely been cast into doubt?

The notion of an attack is another concept that needs to be fitted into this system of classification. In the Carneades system, a proposition can be stated, questioned, assumed or accepted. In Carneades one argument can attack another in basically four ways.

1. It can attack one or more of the premises of the prior argument and show that one or more of them is questionable.

2. It can attack one of these premises and show that one or more of them is not acceptable.
3. It can attack the conclusion by posing a counterargument that shows that the conclusion is questionable.
4. It can attack the conclusion by posing a counterargument that shows that the conclusion is not acceptable.

Is an attack the same thing as a rebuttal? At first, it seems that it is, because an attack on an argument is designed to show that the argument is questionable, that it is not supported by the evidence, or even that the evidence shows that it is untenable. On the other hand, it would seem that it is not, because asking a critical question could perhaps be classified as an attack on an argument, it would not seem quite right to say that asking such a critical question is a rebuttal.

This classification may be borderline, however. Asking a critical question casts doubt on an argument, but is casting doubt on an argument rebutting it? What Carneades has shown is that critical questions matching argumentation schemes are of two different kinds in this regard (Walton and Gordon, 2005). Some critical questions act as rebuttals when they are asked, because unless the proponent of the argument replies appropriately to the question, the argument is defeated. Asking other critical questions does not defeat the original argument unless the question is backed up by some evidence. In this kind of case it does not really seem quite right to describe the asking of the critical question as a rebuttal. The word 'rebuttal' also implies that the attacking is being done by posing another argument, and not merely by asking a question about the original argument, even if it is a critical question that casts doubt on the argument.

In addition to the three basic ways of attacking an argument listed in section 1, we also considered some other ways. One of these ways is to argue that the given argument is not relevant to the ultimate conclusion to be proved in the case at issue. To attack an argument in the fourth way, matters of how the argument was used for some purpose in a context of dialogue needs to be taken into account. Even though the given argument may stand, having repelled all attacks of the first three kinds, its force as argument may be nullified if it is irrelevant. But is this kind of charge a rebuttal? It is not, if it is not an attack on the argument itself, but rather a charge that the argument is not useful for some purpose. A charge of irrelevance is best seen as a procedural objection to the effect that the argument is not useful to resolve the ultimate issue under discussion. To model this kind of procedural objection, we have to look at argumentation as a process, after the manner of Carneades.

## 6. How Carneades Models Relevance

The Carneades system can be used to assist an agent preparing a case by constructing arguments used to prove a claim in a situation where there is an information service that continually provides new information that might be useful for this purpose (Ballnat and Gordon, 2010). The agent only presents his case once the resources provided by the information service have been exhausted. If that has not happened, the agent tries to make his case by asking questions and searching for new information to construct arguments. Then he selects which arguments to put forward in order to prove the goal thesis that he wants to prove. In this system there is a continuous loop as the agent keeps collecting new information from the information service and uses that information to construct new arguments. A simplified version of this process comparable to the figure in (Ballnat and Gordon, 2010, 52) is shown in figure 7.

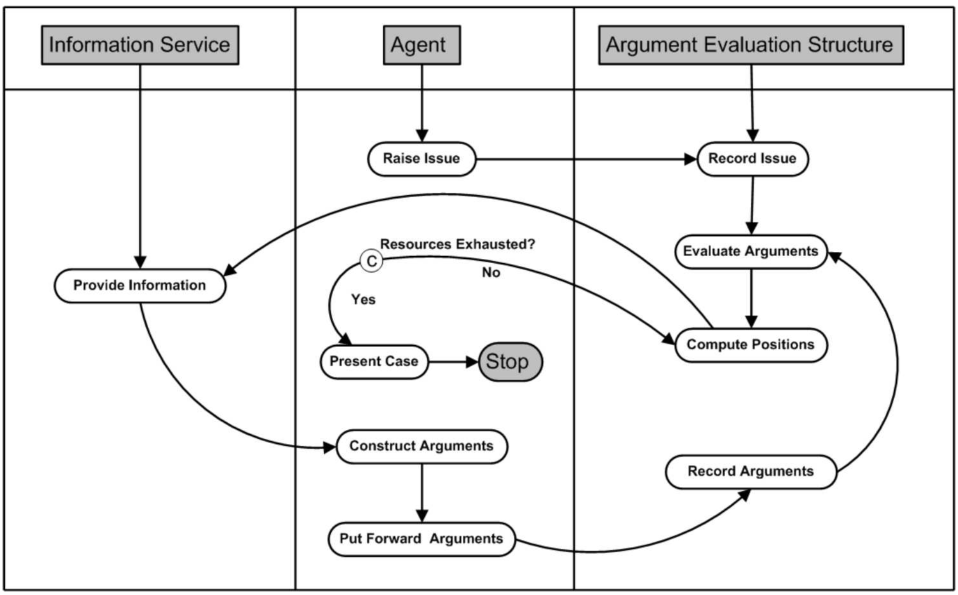



**Figure 7: An Argumentation Process**

Only once these information and argument construction resources are exhausted does the agent either prove his thesis or find that there are insufficient resources to do so. As the agent proceeds through this argumentation process, he tries to find alternative positions to support his argument.

Suppose I want to prove my claim that proposition $A$ is true. What should I do? Should I make a further argument pro $A$? Or should I make another argument con $B$, where $B$ is some proposition that is being used by the opposition to refute $A$? Or should I put forward arguments supporting some premise of one of my previous arguments that were put forward in support of $A$? In other words, what should be my next goal, where a goal is a proposition that a party searches around for to work on next, by looking for arguments *pro* or *con* the proposition he ultimately wants to prove in the dialogue. Carneades is being used here as a device to find which arguments are relevant by telling him which propositions he should choose to work on next, given the information he already has.

As well as providing a method for helping in arguer to determine which arguments are relevant, Carneades can also be used to help in arguer determine which arguments are not relevant. What is presupposed by a claim of relevance is that the given argument is supposed to be used to resolve some unsettled issue in a discussion that is being carried on in the given case. If an argument has no probative value as evidence to prove or disprove the thesis at issue in a particular discussion, it may be dismissed as irrelevant. However, although this attack may knock the argument out of consideration, it is not, strictly speaking, a rebuttal. It should be classified as a procedural objection claiming that the argument under consideration is useless to prove some ultimate claim that the arguer is building a case to prove. On this analysis, the objection to an argument on grounds of relevance is different from the rebuttals and refutations we have been concerned with. Still, it is interesting to see that Carneades has the capability for dealing with claims of relevance and irrelevance because it can model argumentation as a process.

The procedure recommended for seeking some means of refuting or objecting to an argument broadly follows the line of investigation in the paper. It starts out by focusing on refutation in the narrower sense, referring to external and internal refutation, then goes on to means of attack and investigation of an argument offered by argumentation schemes and critical questions. From there, it looks more widely to other kinds of objections that may be procedural in nature, and that may not focus so narrowly on internal or external refutation. As it expands outwards, it takes into account the wider context of an argument, and can do so by viewing argumentation as a process using the Carneades system.

## 7. Classifying Objections, Rebuttals and Refutations

An objection does not necessarily have to be a counter-argument posed against an original argument. It could be merely the asking of a critical question. Even when an objection is a counter-argument posed against an original argument, it does not have to be an argument that the original argument is weak, unsupported or incorrect. It could be a procedural objection, not implying that the argument it is addressed against is incorrect, insufficiently supported by evidence, or even questionable as an argument in itself. Such a procedural objection could merely claim that the argument, even though it might be reasonable enough, or well enough supported in itself, is not appropriate for use in the context of the given discussion. In law for example, an argument might be objected to on the grounds that the evidence it purports to bring forward has been obtained illegally, even though that evidence might otherwise be quite convincing in itself as a rational argument. It follows that an objection is not necessarily a rebuttal or a refutation. The term 'objection' represents a wider category.

There may be a narrower sense of the word 'objection', however, that is used in logic. Govier (1999, 229) considers an objection to be an argument raised against a prior argument. Hence a question is not an objection: "On this view, a question purely considered as such does not itself constitute an objection". On her account, an objection can be directed in one of two ways. The objection can claim that there is something wrong with the conclusion, or it can claim that there is something wrong with the argument. But these are not the only possibilities. She classifies five types of objections (231), depending on what the objection is specifically raised against: (1) against the conclusion, (2) against the argument in support of the conclusion, (3) against the arguer, (4) against the arguer's qualifications, personal characteristics or circumstances, or (5) against the way the argument or conclusion was expressed. It is interesting to note that some of these categories of objection may correspond to or overlap with types of arguments associated with some of the traditional informal fallacies. The third category and two parts of the fourth may correspond to the *ad hominem* type of argument while the first part of the fourth may correspond to a common type of attack on arguments from expert opinion.

A different way of classifying objections to an argument has been put forward by Krabbe (2007, 55–57) who lists seven ways an opponent can critically react to a proponent's expressed argument. (1) A request for clarification, explanation or elucidation may contain an implicit criticism that the argument was not clearly expressed to start with. (2) A challenge to an argument comprises an expression of critical doubt about whether a reason supports the argument. (3) A bound challenge raises a more specific doubtful point that offers some reason for entertaining doubt. (4) An exposure of a flaw poses a negative evaluation of an argument and requests further amplification. (5) Rejection is a kind of critical reaction by an opponent who may not deny that the proponent's argument is reasonable, but takes up an opposite point of view. (6) A charge of fallacy criticizes the contribution of the proponent by claiming he or she has violated some rule of fair procedure. (7) A personal attack is a common kind of critical reaction that provides a means of defense against unreasonable moves by one's opponent. Krabbe (2007, 57) suggests that these critical reactions can properly be called objections, because they expresses dissatisfaction with an argument presented by a proponent. However, Krabbe (2007, 57) writes that to speak of a request for clarification or a pure challenge as an objection would be an overstatement, because objections presuppose a negative evaluation, whereas these other two types of reaction precede evaluation.

There are differences between these two views on what an objection is. Govier (1999, 229), requires that an objection be an argument when she wrote, "An objection is an argument, a consideration put forward, alleged to show either that there is something wrong with the conclusion in question or that there is something wrong with the argument put forward in its favor". Krabbe does hold the view that an objection has to be an argument. Ralph Johnson, in an unpublished manuscript shown to the author, has advocated the view an objection is a response to an argument that can be in the form of a question or a statement, and does not have to be an argument. I will take it that objection is a wider category than rebuttal, so that while putting forward a rebuttal is making an objection in some instances, there are also instances in which an objection to an argument should not be classified as a rebuttal.

The notion of a challenge is well known in argumentation. In his Why-Because System with Questions, Hamblin (1970, chapter 8), has a locution 'Why $A$?' that is a challenge or request made to the hearer to provide a justification (an argument) for the statement $A$ queried. But what is a challenge to an argument (as opposed to a statement)? Most likely it would seem to be a critical question. But there could be other sorts of argument challenge, for example such a challenge could be a procedural objection that the argument is irrelevant.

Following the line of this paper, the notion of a rebuttal can be defined as follows. A rebuttal requires three things. First, it requires a prior argument that it is directed against. Second, the rebuttal itself is an argument that is directed against this prior argument. Third, it is directed against the prior argument in order to show that it is open to doubt or not acceptable.

A rebuttal is one of a pair of arguments, where the two arguments are ordered, logically rather than temporally, so that the one precedes the other, and so that the second one is directed against the first one. What does "directed against" mean? One argument can have another argument as its target. The one can be meant to support the other, or can be meant to attack the other, or the two arguments can be independent of each other. But something more is meant here. What seems to be implied is that a rebuttal is an argument directed against another argument to show that the first argument is somehow defective. To rebut an argument is to try to show that the argument is questionable, or not supported by the evidence, or even that the evidence shows that it is untenable.

Is a refutation the same as a rebuttal? One way to define the relationship between these two terms strongly suggested by our discussion of how Carneades handles the type of argument configuration shown in figure 3 would be to say that a refutation is a successful rebuttal. On this way of defining the two terms, a rebuttal is aimed to show that the argument it is directed against is questionable or untenable. A refutation is a rebuttal that is successful in carrying out its aim. A refutation is a counterargument that is not only posed against a prior argument, but weighs in more strongly when evaluated against the prior argument so that it reverses the conclusion of the prior argument. So defined, the one term would seem to be a subspecies of the other. A refutation is a species of rebuttal that shows that the argument it is aimed at is untenable. When an argument you have put forward is refuted, it has to be given up. If the argument is confronted with a rebuttal, you may or may not have to give it up. Only if the rebuttal is a refutation do you have to give it up. The same point can be made about attack. Attack does not imply defeat.

The term challenge is widely used in formal dialogue systems. As noted above, Hamblin has a locution 'Why $A$?', called a challenge, in his Why-Because System with Questions. To respond appropriately the hearer is expected to provide premises that the challenger is committed to already, or can be brought to concede at future moves), and $A$ is supposed to be a conclusion implied by these premises according to the inference rules in the system. A challenge, in this sense, is not an argument. It is a speech act that requests some evidence to support a claim made by the other party. As the distinction between assumptions and exceptions made in Carneades shows, some critical questions are merely challenges, whereas other critical questions, although they have the speech act format of a challenge, defeat the other party's argument unless she comes forward with some evidence to support her argument.

The classification tree shown in figure 8 offers a way of clarifying these terms.

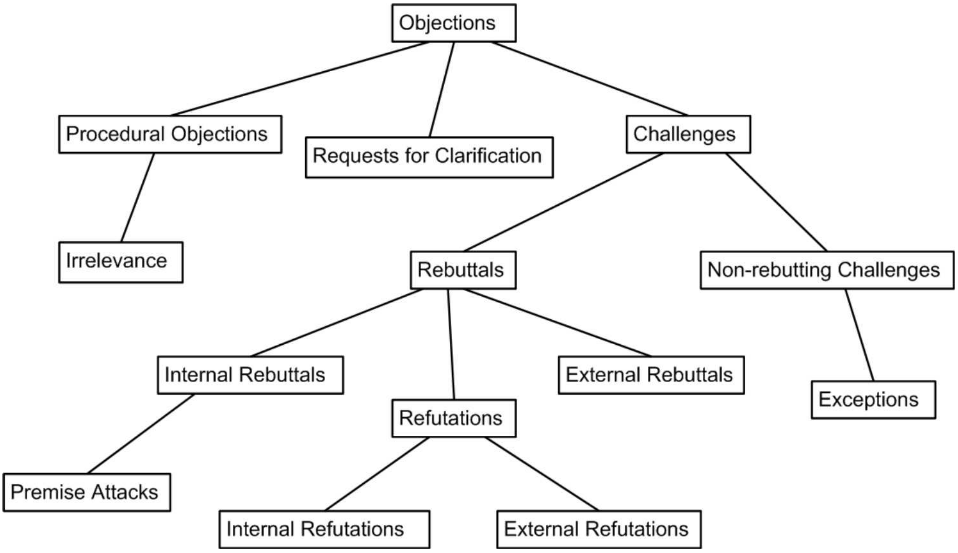



**Figure 8: Classification Tree for Species of Objections**

Objection is taken to be a wide category that includes procedural objections, and many kinds of attacks that should not, strictly speaking, be called rebuttals. An objection of irrelevance is shown as an example of a procedural objection. An objection does not have to be a rebuttal even though it is comparable to a rebuttal in that it assumes that there is something negative about an original argument, or move in argumentation, that needs to be responded to, called into question and corrected. The classification tree in figure 8 incorporates the notion of a challenge. A challenge is defined after the manner of Krabbe as a species of objection that comprises an expression of critical doubt about whether a reason supports the argument that is challenged. However, this way of defining the notion of challenge makes it appear to be very close to a Pollock-style undercutter, a species of argument attack modeled as an exception in Carneades. Figure 8 clarifies the notion of the challenge by classifying the Pollock-style undercutter as an exception, using the term and its Carneades meaning. Exceptions are classified as critical questions that need to be backed up by evidence before they defeat the argument they are directed against. The classification tree shown in figure 8 also incorporates the distinction between an internal refutation or rebuttal and an external one. Hence it is a comprehensive classification scheme that includes all the species of objections analyzed in the paper.

A rebuttal is a species of objection. A refutation is a species of rebuttal that is successful in knocking down the argument it was directed against. A *rebuttal* is an argument directed against another argument to show that the first argument is somehow defective. An *attack*, in the sense of the word as used in the field of argumentation, is an argument directed against another argument to show that the first argument is somehow defective. In other words, for purposes of argumentation study, the words 'rebuttal' and 'attack' can be taken as equivalent.

To rebut an argument is to try to show that the argument is questionable, or that it is not supported by the evidence, or even that the evidence shows that it is untenable. A rebuttal can attack a premise of the original argument, it can attack the conclusion, or it can act as an undercutter that attacks the inference from the premises to the conclusion. How it does this, as illustrated by Pollock's red light example and the Tweety example, is by finding an exception to a general rule that is the warrant of a defensible argument. A *refutation* is a species of rebuttal that shows that the argument it is aimed at is unacceptable. It could be called a knock-down counter-argument. When an argument you have put forward is confronted with a refutation, it has to be given up. Both rebuttals and refutations can be external or internal.

## 8. Conclusion

The practical argument attack and refutation procedure derived from the analysis in this paper has seven steps. The procedure can be applied using these seven steps.

1. Look for a refutation in the sense described in section 2. If you have a counter-argument that can be used to prove the opposite of the conclusion claimed in the original argument, go for an external refutation.

2. Alternatively, if this seems to be a better route of attack, go for an internal refutation.

3. The first step in seeking a suitable internal refutation is to see if the argument you are trying to attack fits a known argumentation scheme. The list of the most basic types of arguments that have argumentation schemes are the following: argument from position to know, argument from witness testimony, argument from expert opinion, argument from analogy, argument from verbal classification, argument from rule, argument from precedent, practical reasoning, value-based practical reasoning, argument from appearances (perception), argument from ignorance, argument from consequences (positive or negative), argument from popular opinion, argument from commitment, direct *ad hominem* argument (personal attack), circumstantial *ad hominem* argument, argument

from bias, argument from correlation to cause, argument from evidence to a hypothesis, abductive reasoning, argument from waste, and slippery slope argument.

4. If the argument fits a scheme that can be identified, look at the critical questions matching the scheme, and see which question is most appropriate.

5. In the Carneades model critical questions are represented as different kinds of premises. If the premise you choose to attack is either an ordinary premise or an assumption, simply question it.

6. If it is an exception, question it only if you have the evidence required to back it up.

7. If none of this procedure so far has come up with a good result, go on to look for some procedural objection, like questioning whether the argument is relevant.

Throughout the main part of paper the narrower concern has been with the concept of refutation illustrated by Goodwin's example that we began with. But later there was a move to considering other kinds of objections that, it was argued, do not fit this narrower model. The list of objections provided by Krabbe gives a good idea of what some of these objections are, but there is no reason to think that this list is complete. Some of the objections correspond to informal fallacies of the kind well known in the argumentation literature. Objecting to an argument on the grounds that it is circular and therefore begs the question is an example. The task of studying and classifying additional kinds of objections to an argument associated with fallacies is a project for future research.

## References

Ballnat, S. and Gordon, T. F. (2010). Goal Selection in Argumentation Processes, *Computational Models of Argument: Proceedings of COMMA 2010*, ed. P. Baroni, F. Cerutti, M. Giacomin and G. R. Simari, Amsterdam: IOS Press, 51–62.

Dung, P. M. (1995). On the Acceptability of Arguments and Its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-person Games, *Artificial Intelligence*, 77, 321–357.

Goodwin, J. (2010). How to Refute an Argument, accessed November 26, 2010 at this site: http://www.public.iastate.edu/~goodwin/spcom322/refute.pdf

Gordon, Thomas F. (2010). An Overview of the Carneades Argumentation Support System, *Dialectics, Dialogue and Argumentation*, ed. C. Reed and C. W. Tindale, London: College Publications, 145–156.

Gordon Thomas F. and Walton, D. (2006). The Carneades Argumentation Framework, *Computational Models of Argument: Proceedings of COMMA 2006*, ed. P. E. Dunne and T. J. M. Bench-Capon. Amsterdam: IOS Press, 195–207.

Gordon, T. F., & Walton, D. (2009). Proof Burdens and Standards, ed I. Rahwan & G. Simari *Argumentation in Artificial Intelligence*, Berlin, Germany: Springer-Verlag, 239–260.

Gordon, T. F., Prakken, H. and Walton, D. (2007). The Carneades Model of Argument and Burden of Proof, *Artificial Intelligence*, 171, 875–896.

Govier, T. (1999). *The Philosophy of Argument*. Newport News, Virginia: Vale Press.

Govier, T. (2006). *The Philosophy of Argument*, Newport News, Virginia, Vale Press, 1999.

Hamblin, C. L. (1970). *Fallacies*. London: Methuen.

Hitchcock, D. and B. Verheij (eds) (2006), *Arguing on the Toulmin Model: New Essays in Argument Analysis and Evaluation*. Dordrecht: 2006.

Krabbe, E. C. W. (2007). Nothing but Objections!, *Reason Reclaimed*, ed. H. V. Hansen and

Pollock, J. (1995). *Cognitive Carpentry*, Cambridge, Mass., MIT Press.

Prakken, H. (2003). Logical Dialectics: The Missing Link Between Deductivism and Pragma-Dialectics, *Proceedings of the Fifth Conference of the International Society for the Study of Argumentation*, ed. Frans H. van Eemeren at al., Amsterdam: SicSat, 857–860.

Prakken, H. (2010). On the Nature of Argument Schemes, *Dialectics, Dialogue and Argumentation*, ed. C. Reed and C. W. Tindale, London, College Publications.

Toulmin, S. (1952). *The Uses of Argument*, Cambridge, Cambridge University Press.

Verheij, B. (2009). The Toulmin Argument Model in Artificial Intelligence, *Argumentation in Artificial Intelligence*, ed. I. Rahwan and G. Simari, Berlin, Springer, 2009, 219–238.

Walton, D. (1997). *Appeal to Expert Opinion*, University Park, Pa: Penn State Press.

Walton, D. (2006). Argument from Appearance: A New Argumentation Scheme, *Logique et Analyse*, 195, 2006, 319–340.

Walton, D. and Godden, D. M. (2005). The Nature and Status of Critical Questions in Argumentation Schemes, *The Uses of Argument: Proceedings of a Conference at McMaster University*, ed. D. Hitchcock, Hamilton, Ontario: OSSA, 476–484.

Walton, D. and Gordon, T. F. (2005). Critical Questions in Computational Models of Legal Argument. *Argumentation in Artificial Intelligence and Law*, IAAIL Workshop Series, ed. Dunne, P. E. and T. J. M. Bench–Capon, Nijmegen: Wolf Legal Publishers, 103–111.

Walton, D., Reed, C. and Macagno, F. (2008). *Argumentation Schemes*, Cambridge: Cambridge University Press.

Douglas Walton
Centre for Research in Reasoning,
Argumentation and Rhetoric (CRRAR)
University of Windsor
2500 University Ave. W., Windsor, Ontario N9B 3Y1, Canada
dwalton@uwindsor.ca

**Paweł Łoziński**
Warsaw University of Technology

# AN ALGORITHM FOR INCREMENTAL ARGUMENTATION ANALYSIS IN CARNEADES

**Abstract**: *Carneades* is an interactive application for argument construction, evaluation and visualization, integrating an knowledge-based inference engine and an argument mapping tool. Given the argument sources and the goal of argumentation process, *Carneades* conducts a resource-limited search for arguments for and against the given goal. The result of the search is an argumentation graph which can then be visualized and analized by the application user, e.g. an expert in the legal domain. This article presents a different, incremental approach to the exploration of argumentation space with a search algorithm using heuristics and search constraints for choosing the exploration paths. The article describes a motivation for such an approach, construction and implementation of the algorithm together with its comparison to argument construction in *Carneades*.

**Keywords**: argumentation theory, logic programming, reasoning, Carneades

## Introduction

*Carneades* is an interactive application for argument construction, evaluation and visualization, integrating an knowledge-based inference engine and an argument mapping tool. As described in [3], *Carneades* is designed to handle multiple sources of arguments including ontologies, rules, cases and testimonial evidence.

Given the argument sources and the goal of the argumentation process, *Carneades* conducts a resource-limited search for arguments for and against the given goal. The result of the search is an argumentation graph which can then be visualized and analized by the application user, e.g. an expert in the legal domain. All the statements in the graph are labeled to indicate whether the statement and it's compilment are acceptable. This method gives the user a full (to the extent of resources avaliable during the search) picture of the subject of discourse, which is both an advantage and a potential source of problems. The advantage is that the user can see and analize the whole picture. However, as it can be seen from argumentation example presented in [4], the graph of arguments that is presented to the

ISBN 978–83–7431–305–6 ISSN 0860-150X 

user can be very large, even for a relatively small argumentation case. This makes it difficult for the user to grasp and draw conclusions from. In the software licensing example presented in [4], the argumentation graph for goal *Exists copyright license that may be used by carneades engine* exceeds 100 nodes. Also, the time needed to construct the graph using e.g. arguments generated from ontology, which involves quering the OWL reasoner, can be substantial.

This article presents a different approach to the exploration of argumentation space with a search algorithm using heuristics and search constraints for choosing the exploration paths. The algorithm aims to find a minimal argumentation graph that allows for determining the acceptability of the given goal and to find it in minimal number of steps. In this approach, the analysis of an argumentation case takes an incremental form. After viewing one portion of information avaliable in the argumentation graph, the user may want to query the system with a different goal, using the obtained knowledge.

The rest of the paper is organized as follows. Section 1 briefly describes *Carneades* as defined in [6]. Section two introduces a reformulation of *Carneades* in terms of inference rules. It is important to note that *Carneades* has a notion of arguments *from* rules (e.g. legal rules) as opposed to e.g. arguments from ontologies. In this paper the term rule is used in a different meaning, as an inference pattern and in this sense an argument may be viewed as a rule. Section three presents a general approach to constructing an argumentation space search algorithm. In section four the RPA* search algorithm is presented. In section five the implementation of the algorithm is described. Section six contains a brief comparison of the argumentation graph generation using latest implementation of *Carneades* and $RPA*$. The paper concludes with a brief summary and description of future work on the subject.

## 1. *Carneades*

**Definition 1** (Statements)

Let $(\mathcal{L}, =, complement)$ be a structure where $\mathcal{L}$ denotes a set of declarative statements in some language, "=" is an equality relation modeled as a function of type $\mathcal{L} \times \mathcal{L} \rightarrow boolean$, and $complement : \mathcal{L} \rightarrow \mathcal{L}$ is a function mapping a statement to its logical complement. If $s$ is a statement, the complement of $s$ is denoted $\bar{s}$.

**Definition 2** (Premises)

Let $\mathcal{P}_{\mathcal{L}}$ denote the set of premises. There are the following types of premises: (1) If $s \in \mathcal{L}$, then $\diamond s$, called *ordinary premise*, is a premise. (2) If $s \in \mathcal{L}$, then $\bullet s$, called *assumption*, is a premise. (3) If $s \in \mathcal{L}$, then $\circ s$, called *exception*, is a premise. (4) Nothing else is a premise.

**Definition 3** (Arguments)

An *argument* is a tuple $(c, d, P)$, where $c \in \mathcal{L}$, $d \in \{pro, con\}$ and $P \in 2^{\mathcal{P}_{\mathcal{L}}}$.

The key notion in *Carneades* framework is an argument graph, on the bases of which acceptability of statements can be determined.

**Definition 4** (Argument graphs)

An *argument-graph* is a labeled, finite, directed, acyclic, bipartite graph, consisting of *argument* nodes and *statement* nodes. The edges link the argument nodes to the statements in the premises and conclusion of each argument. At most one statement node is allowed for each statement $s$ and its complement, $\bar{s}$.

A fragment of an argument graph with one argument node and four statement nodes is shown in 1. Due to the limitation present in this definition, for the sake of graph representation, the notion of *negative premise* is introduced. Premise of every type can be linked to an argument as a negated premise, which is denoted: $\diamond\bar{s}$, $\bullet\bar{s}$, $\circ\bar{s}$.

**Definition 5** (Argument context)

Let $\mathcal{C}$, the *argument context*, be a tuple $(status, ps, >)$, where *status* is a function of type $\mathcal{L} \rightarrow \{stated, questioned, accepted, rejected\}$, *ps* is a function of type $\mathcal{L} \rightarrow \mathcal{PS}$ and "$>$" is a strict partial ordering on arguments. $\mathcal{PS}$ is the set of proof standards. For every statement $s$ and its complement $\bar{s}$, the proof standard assigned to $\bar{s}$ is the complement of the proof standard assigned to $s$ and

- if $status(s) = stated$ then $status(\bar{s}) = stated$,
- if $status(s) = questioned$ then $status(\bar{s}) = questioned$,
- if $status(s) = accepted$ then $status(\bar{s}) = rejected$, and
- if $status(s) = rejected$ then $status(\bar{s}) = accepted$.

Statement is acceptable in the given argument graph if its proof standard is satisfied in this graph. In [6] three proof standards are defined (as stated in the article, this list is neither complete nor mandatory, a CAF-based system can have other proof standards defined).

**SE** **(Scintilla of Evidence)** A statement meets this standard iff it is supported by at least one defensible pro argument.

**BA** **(Best Argument)** A statement meets this standard iff it is supported by some defensible pro argument with priority over all defensible con arguments.

**DV** **(Dialectical Validity)** A statement meets this standard iff it is supported by at least one defensible pro argument and none of its con arguments are defensible.

The *complement* of a proof standard $\sigma$, denoted $\bar{\sigma}$, is a standard which results from switching the roles of pro and con arguments in the definition of $\sigma$.

An argument is *defensible* if all of its premises *hold*. Holding of a premise depends firstly on its type and status of its statement: premise $\diamond s$ holds if $status(s) = accepted$ and doesn't hold if $status(s) = rejected$; $\bullet s$ holds if $status(s) \in \{accepted, stated\}$ and doesn't hold if $status(s) = rejected$; $\circ s$ holds if $status(s) = rejected$ and doesn't hold if $status(s) = accepted$. Secondly, in remaining cases, premise $\diamond s$ or $\bullet s$ holds if statement $s$ is acceptable, premise $\circ s$, holds if statement $s$ is not acceptable.

## Rule-based version of *Carneades*

The idea underlying rule-based version of *Carneades* is quite intuitive, i.e. an argument shown in figure 1 can be replaced with an inference rule that uses statements $p$, $q$ and $r$ to conclude $s$. Due to space limitations, we present in this section only a shortened definition, based much on intuition and analogy to the original *Carneades* model. The notion of statement's *status* in CAF can be translated to logical value of a statement, and the notion of *acceptability*, similarly as in other approaches to argument-based logic programming, replaces the notion of truthfulness of a statement. More formally speaking:

**Definition 6** (RCAF)

Logic *RCAF* (rule version of CAF) is and ordered triple $(\mathcal{L}, \mathcal{S}, \mathcal{IM})$ where $\mathcal{L}$ is a simplified version of FOL language, $\mathcal{S}$ is semantics and $\mathcal{IM}$ is an inference mechanism.

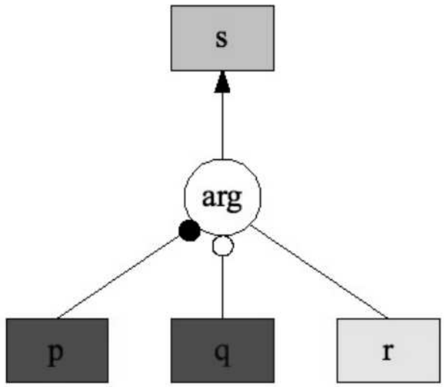



**Figure 1. An argument in *Carneades* and a corresponding rule**

**Definition 7** (Language)

*Language* $\mathcal{L}$ is a traditional FOL language limited to constants, variables, predicates, negation and existential quantifier, e.g. $\neg\exists_x Likes(John, x)$.

**Definition 8** (Semantics)

*Semantics* $\mathcal{S}$ is a standard semantics for $\mathcal{L}$, with difference in the set of logical values, which has four elements: $\{accepted, rejected, stated, questioned\}$. The values of propositions and their negations correspond to statuses of statements and their complements in CAF.

**Definition 9** (Inference mechanism)

*Inference mechanism* $\mathcal{IM}$ consists of three elements: (i) Function $ps$ : $\mathcal{L} \to \mathcal{PS}$, which assigns every proposition in $\mathcal{L}$ its proof standard. (ii) Set $\mathcal{R}$ of defined inference rules. (iii) Relation "$>$" $\subseteq \mathcal{R} \times \mathcal{R}$ of strict partial ordering among rules of inference.

**Definition 10** (Inference rule)

*Inference rule* $\mathfrak{r} \in \mathcal{R}$ is a structure of two possible types $p_1, \ldots, p_n \xrightarrow{pro} c$ or $p_1, \ldots, p_n \xrightarrow{con} c$, where $p_1, \ldots, p_n$ are premises and $c \in \mathcal{L}$ is the conclusion. Rules of the first type are called *pro* rules, those of second type are called *con* rules. We say that rule $\mathfrak{r}$ *supports* conclusion $c$ (regardless of the type of the rule). Type of rule $\mathfrak{r}$ is denoted with $T(\mathfrak{r})$, "$\rightarrow$" denotes a rule of either type.

Definition of rule's premises is analogous to def. 2, the proposition of the premise is sometimes refered to as the premise itself, type of the premise $p$

is denoted $T(p)$. Defensibility of rules and acceptability of propositions is analogous to defensibility of arguments and acceptability of statements in CAF. if its proof stanard is satisfied by ground rules supporting it.

**Definition 11** (Satisfaction of proof standard)

We say that the proof standard of ground proposition $s$ is *satisfied* if the set of ground rules supporting $s$ satisfies the condition given in the proof standard's definition (e.g. for $DV$: there exists a *pro* rule that supports $s$ and there is no *con* rule that supports $s$).

The condition for satisfaction of complement proof standard is generated from the proof standard's definition be switching types of rules and adding negation to the checked proposition.

**Example 1**

Let proposition $s =$"Cat is black" and $ps(s) = SE$. The proposition $s$ is acceptable iff $SE(s) = true$, that is, iff there exists at least one *pro* rule supporting $s$. The proposition $\neg s =$"Cat is not black" is acceptable iff $\overline{SE}(\neg s) = true$, that is, iff there exists at least one *con* rule supporting $\neg\neg s$, that is, one *con* rule supporting $s$.

## 3. Logic programming in RCAF

Proving a ground proposition $s$ in RCAF involves finding such substitution of variables in rules, that the set of all rules in $\mathcal{R}$ supporting $s$ satisfies $ps(s)$. The task seems difficult, because it involves searching for several proofs[1] of $s$ that together satisfy proof standard of $s$. The problem repeats recursively for premises of rules used to prove $s$.

The solution proposed in this paper is based on the idea presented e.g. in [1], that inference in logic can be modeled as a search problem, which in turn can be solved with one of many "off-shelf" search algorithms. This approach allows a clear formulation of the problem of inference in RCAF and makes relevant all the knowledge gathered in the well researched domain of search in AI.

### 3.1. Inference as search

The formulation of the inference problem in terms of search is based on the idea, that the set of proofs of a given proposition can be regarded as a

[1] In the classical sense of the word.

search space that needs to be explored in order to find a complete, sound proof of the proposition. For a general, formal definition of this search task the reader is referred to [1].

### 3.2 Formulation of the search problem in RCAF

Before moving forward, some details should be established, that — although not related to the particular idea of RCAF logic — are necessary to fully define the search task: (a) *Interpretation of variables.* Variables present in premises of proofs are interpreted as being tied with the existential quantifier (e.g. query $King(x)$ is interpreted as $\exists_x King(x)$ and read "Is there a King?"). (b) *Default values.* Even if all propositions in $KB$ have the status and proof standard assigned, those assignments are not established for intermediate products of reasoning. This problem is here solved simply by the introduction of default values: status *stated* and proof standard $DV$. (c) *Priority of assumptions and exceptions.* Assumptions and exceptions are treated with low priority, that is, as long as the reasoning (e.g. about premise $\bullet R(A)$) can continue, this information is not used to determine holding of the premise.

#### 3.2.1. Search space

Let RCAF = $(\mathcal{L}, \mathcal{S}, \mathcal{IM})$, $KB$ a knowledge base defined in that logic and $q \in \mathcal{L}$ be a ground proposition that is a query asked to the knowledge base with a given proof standard. The query is given value *questioned* and interpreted as an ordinary premise $\diamond q$. Holding of this premise is equivalent to proving $q$.

**Definition 12** (Proof)

*Proof* of $q$ is a directed graph $P = (V^P, E^P)$, where $V^P$ is a set of vertices represented by ground premises. If edge $(p_1, p_2)$ exists in $E^P$, then there exists $\mathfrak{r} \in KB$, which has one premise equal to $p_1$ and conclusion equal to the proposition of $p_2$ (after appropriate variable substitution).

*Subgoal* of proof $P$ is a premise $q' \in V^P$ such, that $deg_{in}(q') = 0$ and the proof standard of $q'$ needs to be checked. Proof is called *complete* iff it has no subgoals, otherwise it is *incomplete*. The only vertex of $P$ that has no outgoing edges is $q$ ($deg_{out}(q) = 0$). Number of subgoals in $P$ is denoted with $\delta(P)$.

**Definition 13** (Proof space)

*Proof space* $\mathcal{P}(q)$ is a set of proofs of $q$. We say that proofs $P_1$ and $P_2$ are *neighbouring* in $\mathcal{P}(q)$ iff $P_2$ is constructed from $P_1$ with one *inference* *step* made using subgoal $q_1 \in V_1^P$. Let $V_1^P = V \cup \{q_i\}_{i=1}^m$, $E_1^P = E \cup E'$, where $E' = \{(q_i, p_i)\}_{i=1}^m$, $\{q_i\}_{i=1}^m$ are subgoals of $P_1$ and $\{p_i\}_{i=1}^m$ are some premises of $P_1$ that are incident with subgoals. Sets $V, E, E'$ may equal $\emptyset$. Two types of inference steps are distinguished:

*Through proposition*: Matching $q_1$ with a proposition $s \in KB$. In this case we switch $q_1$ with a new premise $q_1' = T(q_1)s$: $V_2^P = V \cup \{q_i'\}_{i=1}^m$, $E_2^P = E \cup \{(q_i', p_i)\}_{i=1}^m$.

*Through rule*: Matching $q_1$ with head of rule $\mathfrak{r} \in KB$, where $\mathfrak{r} = r_1, \ldots, r_n \to c$. In this case we switch $q_1$ with a new premise $q_1' = T(q_1)c'$ and add rule body: $V_2^P = V \cup \{q_i'\}_{i=1}^m \cup \{r_i'\}_{i=1}^n$, $E_2^P = E \cup \{(q_i', p_i)\}_{i=1}^m \cup \{(r_i', q_1')\}_{i=1}^n$.

Propositions $c', r_1', \ldots, r_n', q_2', \ldots, q_m'$ denote accordingly: elements of rule $\mathfrak{r}$ and subgoals of $P_1$ with applied substitution of variables. In case of the inference step through rule $\mathfrak{r}$ edges from the set $\{(r_i', q_1')\}_{i=1}^n$ are labeled with $\mathfrak{r}$. Proof $P_1$ is called a *predecessor* of $P_2$, which in turn is called a *successor* of $P_1$.

**Example 2**

Let $q = P(x)$ and the knowledge base is

$$KB = \{P(A) : \mathit{accepted}, R(B) : \mathit{accepted}\} \cup \{[\diamond Q(y,z), \bullet R(z)] \xrightarrow{pro} P(z)\}$$

The incomplete proof $P_0 = (\{\diamond P(x)\}, \emptyset)$ has two neighbouring proofs (successors of $P_0$), $P_1$ and $P_2$, which in turn has successor $P_3$:

$$P_1 = (\{\diamond P(A)\}, \emptyset),$$
$$P_2 = (\{\diamond P(x), \bullet R(x), \diamond Q(y,x)\}, \{(\diamond Q(y,x), \diamond P(x)), (\bullet R(x), \diamond P(x))\}),$$
$$P_3 = (\{\diamond P(x), \bullet R(B), \diamond Q(y,B)\}, \{(\diamond Q(y,B), \diamond P(x)), (\bullet R(B), \diamond P(x))\}).$$

$P_1$ is a complete proof, $P_2$ has two subgoals: $Q(y,x)$ i $R(x)$, $P_3$ has one subgoal: $Q(y,B)$ (which is read: $\exists_y Q(y,B)$). The proof space can be interpreted as an (undirected in this case) graph:

$$\mathcal{P}(q) = (\{P_0, P_1, P_2, P_3\}, \{\{P_0, P_1\}, \{P_0, P_2\}, \{P_2, P_3\}\}).$$

The default starting point for search in the proof space $\mathcal{P}(q)$ equals $P_0 = (\{\diamond q\}, \emptyset)$. As shown in the example 2, proof space $\mathcal{P}(q)$ can be represented as a graph, whose vertices are proofs and edges represent the neighbouring relation. The edges can be given a direction, e.g. form the predecessor to the successor.

### 3.2.2. Search task

As it was mentioned at the beginning of section 3, the goal of search in the RCAF's proof space – unlike in the standard case – is not to find a single complete proof of proposition $q$, but rather to find a set of complete proofs $\mathcal{P}^t = \{P_1^t, \ldots, P_n^t\}$ that together satisfy proof standard $ps(q)$. Moreover, $\mathcal{P}^t$ should be maximal in the sense, that proof space cannot contain any other complete proof $P^t$ such that $\mathcal{P}^t \cup \{P^t\}$ would no longer satisfy $ps(q)$.

Finding the set $\mathcal{P}^t$ involves repeated searching for its elements in $\mathcal{P}(q)$. The found proofs must not require contradictory substitutions, so after finding the complete proof $P_1^t$, the substitution $\theta^t$ of variables used in the proof must be stored. Next, the search should continue from the beginning with $\theta^t$ treated as a constraint: substitutions used in the explored proofs *must not* be contradictory with $\theta^t$. In case of finding the next complete proof $P_2^t$, substitution $\theta_2^t$ should be composed[2] with the previous one: $\theta^t := \theta^t \cup \theta_2^t$, an so on with subsequent proofs.

The search task ends if one of two possible situations occurs:

1. the maximal set $\mathcal{P}^t$ satisfying $ps(q)$ was found (and so $q$ is acceptable),
2. the set $\mathcal{P}^t$ is not a subset of $\mathcal{P}(q)$.

If the constructed set $\mathcal{P}^t$ cannot satisfy $ps(q)$ (e.g. just found proof $P_n^t$ uses rule $A \xrightarrow{con} q$, when $ps(q) = DV$), then the following steps should be taken: (a) the constructed substitution $\theta^t$ should be stored in a different record: $\theta^f := \theta^t$; (b) search results should be cleared: $\mathcal{P}^t := \emptyset$ and $\theta^t := \emptyset$; (c) the search should continue from the beginning with $\theta^f$ treated as a constraint: substitutions used in the explored proofs *must* be contradictory with $\theta^f$. If the situation will repeat, that is, subsequent "false paths" will be found, then every subsequent "forbidden substitution" should be stored: $\Theta^f = \{\theta^f, \theta_1^f, \ldots\}$. Substitution used in the explored proofs must be contradictory with each element of $\Theta^f$.

Finally, if a given subgoal $q'$ is already acceptable in $\mathcal{P}^t$, then it can be memorized (e.g. added to set $Q$) and omitted in further search. The $Q$ set should be cleared everytime $\mathcal{P}^t$ is cleared.

To summarize, given this representation, the inference task in RCAF reduces to the task of repeated search with two types of constraints:

1. non-contradiction with hitherto found proofs,
2. contradiction with proofs leading to refutation of $q$.

[2] See [9, p. 288] for details on substitution composition.

## 4. The RPA* search algorithm

Space $\mathcal{P}(q)$ has a natural starting point for search in $P_0 = (\{\diamond q\}, \emptyset)$ and it contains some information that allows guessing in which direction the complete proof should be looked for (e.g. number of subgoals, proof standards of the subgoals, etc.). Therefore a natural candidate for a search algorithm is $A^*$ (see e.g. [7]).

In its basic version, $A^*$ chooses among possible points of exploration the point $n$, such that $f(n)$ is minimum, and expands it. For every successor $n'$ of $n$ the value $f(n')$ is calculated: $f(n') = g(n') + h(n')$. Value $g(n') = g(n) + c(n, n')$ is a cost of reaching point $n'$ and the value $h(n')$ is an estimated cost of reaching the goal of the search from this point.

While reasoning in RCAF, the function $f$ must carry out two tasks:

**Task 1.** Analogously to the classical case, it should estimate the cost of reaching a complete proof.

**Task 2.** The function should direct the search in such a way, that the algorithm could verify as quickly as possible if the constructed set $\mathcal{P}^t$ can satisfy $ps(q)$.

Accomplishing task 1 allows to find a complete proof in a single search fast, while accomplishing task 2 minimizes the total number of searches needed to construct $\mathcal{P}^t$. This task can be carried out by first checking those proofs, which can guarantee satisfaction or non-satisfaction of the proof standard.

As task 2 exceeds the classical application of $A^*$ algorithm, function $f : \mathcal{P}(q) \to \Re$ will be responsible only for task 1. Let $P$ be a proof, then:

$$f(P) = |V(P)| + \delta(P), \tag{1}$$

where $\delta(P)$ (the $h$ function) is the number of subgoals in $P$. The heuristic assumption is made, that the estimated future cost is equal to the number of subgoals of the proof. So defined function $h$ is not *admissible* (see [7, p. 77]) because, in the optimistic case, substitution in one inference step can prove even all subgoals of the proof being extended.

For addressing task 2, a different function, $p : \mathcal{P}(q) \to \Re$, is defined. It is used to prioritize points of space exploration. $A^*$ algorithm is modified as to first choose points with the highest value of $p$, and among them, the one which has lowest value of $f$.

**Definition 14** ($PA^*$ algorithm)

Let $N$ denote the set of points of search space exploration. $PA^*$ (*prioritized $A^*$) algorithm* is a modification of $A^*$ where the choice of the next point of exploration is limited to set $\{n \in N : \forall_{m \in N} p(n) \geq p(m)\}$, where $p : \mathcal{P}(q) \rightarrow \Re$ is a function that assigns priorities to points of search space exploration.

To construct the $p$ function it is necessary to look more closely at the differences between different successors of a given proof. These differences form a hierarchy:

1. successors can use different subgoals of their predecessors;
2. for the same subgoal $q'$, successors can differ in the type of inference step (through proposition or through rule),
3. for the same type of inference step, successors can use different elements of knowledge base (different rules or different propositions).

Starting from the top level of this hierarchy, for realization of task 2 the difference in used subgoal is irrelevant, whereas the type of inference clearly is: it is generally faster to prove subgoal by matching it with a proposition in knowledge base. If so, the $p$ function should prefer inference through proposition over inference through rule. On the bottom level: (a) among different propositions, the algorithm should prefer those which end proof of the subgoal (those which have value *acceptable* or *rejected*, for assumptions also *stated*); (b) among different rules four levels of priority can be distinguished:

- rules whose defensibility is a *sufficient condition* for not holding of premise $q'$ (level 3);
- rules whose defensibility is a *sufficient condition* for holding of $q'$ (level 2);
- rules whose defensibility is a *necessary condition* for holding or not holding of premise $q'$ (level 1);
- other rules (level 0).

Rules that definitively prevent holding of the premise have the highest priority, because defensibility of such rule for any subgoal means automatic failure of the chosen exploration path. Depending only on the proof standard, the assignment of priorities to rules is as follows:

**SE:** *pro* rules: level 2, *con* rules: level 0;

**BA:** maximal w.r.t. ">" rules *con*: level 3, maximal w.r.t. ">" rules *pro*: level 1, other rules: level 0;

**DV:** *con* rules: level 3, *pro* rules: level 1.

For complementary proof standards the above mentioned types of rules should be switched. For $T(q') = \circ$, the above assignments of level 2 and 3 to proof standards should be switched, because an exception holds if its proof standard is *not* satisfied. Of course, it is possible to differentiate rules priorities with other criteria (e.g. number of premises), it is also possible to construct a more refined rule hierarchy for the $BA$ proof standard; these can be regarded as optimization steps, that are not crucial for the concept of $PA^*$ algorithm.

After defining levels of priority, it is possible to formally define the $p$ function. First, $p(P_0) = 0$, where $P_0 = (\{\diamond q\}, \emptyset)$ is the starting point of search. Let $P'$ be a proof, which was constructed from $P$ (is its successor) by matching subgoal $q_P$ with (i) proposition $s$, (ii) head of rule $\mathfrak{r}$. Function $p$ is defined as follows.

$$p(P') = \begin{cases} p(P) + 4 & \text{in case (i)}, \\ p(P) + prir(q_P, \mathfrak{r}) & \text{in case (ii)}. \end{cases} \quad (2)$$

Help function $prir : \mathcal{L} \times \mathcal{R} \rightarrow \Re$ equals ($\mathcal{R}^{q_P}_{max}$ denotes the set of maximal (w.r.t ">") rules supporting $q_P$):

- If $ps(q_P) = SE$ then

$$prir(q_P, \mathfrak{r}) = \begin{cases} 3 & \text{if } T(q_P) = \circ \wedge T(\mathfrak{r}) = \mathit{pro}, \\ 2 & \text{if } T(q_P) \neq \circ \wedge T(\mathfrak{r}) = \mathit{pro}, \\ 0 & \text{if } T(\mathfrak{r}) = \mathit{con}. \end{cases}$$

- If $ps(q_P) = BA$ then

$$prir(q_P, \mathfrak{r}) = \begin{cases} 3 & \text{if } T(q_P) \neq \circ \wedge T(\mathfrak{r}) = \mathit{con} \wedge \mathfrak{r} \in \mathcal{R}^{q_P}_{max}, \\ 2 & \text{if } T(q_P) = \circ \wedge T(\mathfrak{r}) = \mathit{con} \wedge \mathfrak{r} \in \mathcal{R}^{q_P}_{max}, \\ 1 & \text{if } T(\mathfrak{r}) = \mathit{pro} \wedge \mathfrak{r} \in \mathcal{R}^{q_P}_{max}, \\ 0 & \text{otherwise}. \end{cases}$$

- If $ps(q_P) = DV$ then

$$prir(q_P, \mathfrak{r}) = \begin{cases} 3 & \text{if } T(q_P) \neq \circ \wedge T(\mathfrak{r}) = \mathit{con}, \\ 2 & \text{if } T(q_P) = \circ \wedge T(\mathfrak{r}) = \mathit{con}, \\ 1 & \text{if } T(\mathfrak{r}) = \mathit{pro}. \end{cases}$$

The proposed algorithm for reasoning in RCAF is $RPA^*$:

**Definition 15** ($RPA^*$ algorithm)

$RPA^*$ (*repeated* $PA^*$) *algorithm* is a repeated execution of $PA^*$ algorithm with function $p$ given by equation 2 and function $f$ given by equation 1. The set of legal points of space exploration is modified according to constraints given in section 3.2.2. Repeated execution of $PA^*$ algorithm ends when one of the stop conditions given in section 3.2.2 is fulfilled. $RPA^*$ concludes that the found $\mathcal{P}^t$ set is maximal if $\mathcal{P}(q)$ contains no points of search space exploration that are not contradictory with $\theta^t$.

All elements of $\mathcal{P}^t$ are directed, acyclic argument graphs with root equal to $q$. A single argumentation graph, which should be the result of argumentation space search can be obtained be summing together elements of $\mathcal{P}^t$.

## 5. Implementation

The above algorithm was implemented in Java language, using JUNG framework (`http://jung.sourceforge.net`) for graph manipulation and visualization. It is avaliable for download at: `http://www.ii.pw.edu.pl/~plozinsk/materialy/rcaf-demo.jar`. The implementation uses a compression method for search space storage which is based on the observation that neighbouring proofs differ only with one inference step, so they can be stored in a differential manner. The program is equipped with a GUI which supports (a) editing the knowledge base (in a textual form); (b) loading text with the knowledge base to the reasoner; (c) asking queries with specified proof standard. A GUI shows visualization of the elements of $\mathcal{P}^t$ mapped into one directed graph. The colours of propositions in the graph follow the street lights metaphor: green represents status *accepted*, yellow is for *questioned*, red for *rejected* and additionally grey for *stated*. Green is also used to mark *pro* rules, while *con* rules are marked with read.

The format for textual editing of knowladge base is as follows. Predicates and constants must begin with uppercase, variables must begin with lowercase. Every proposition is entered in separate line and has two optional parameters: `s` used to assign the proposition's status and `p` to assign the proof standard. Default values are `s=STATED` and `p=DV`. Rules have optional labels that can be used in specifying the partial ordering, e.g. $r1 < r2$. Rules of type *pro* are denoted with "`->`", *con* with "`-<`". Assumptions are marked with "`+`", and exceptions with "`-`". Negation is marked with "`!`". An example knowledge base will be given in the next section.

## 6. Short comparison

The results of argumentation space search with *Carneades*'s abductive construction of arguments and $RPA*$ algorithm will be compared using rather abstract, but concise example. For the purpose of the comparison, the latest implementation of *Carneades* is used. The system is implemented in *Clojure* functional programming language, and avaliable for download form `http://carneades.berlios.de/`. Let us consider a knowledge base defined in *Carneades* as simply as possible:

```
(def abstract-rb
    (rulebase
      (rule r1
            (if (B ?x) (A ?x)))
      (rule r2
            (if (and (D ?x) (E ?x ?y)) (B ?x)))
      (rule r3
            (if (I ?x) (not (B ?x))))
      (rule r4
            (if (and (D ?x) (F ?x ?y)) (I ?x)))
      (rule r5
            (if (C ?x) (not (A ?x))))
      (rule r6
            (if (and (G ?x) (E ?x ?y)) (C ?x)))
      ))
```

Using this knowledge base for the only source of arguments, we ask if statement $A(X)$ is acceptable with proof standard $DV$. First, we accept the following facts: $D(X)$, $E(X,Y)$ and $F(X,Y)$. Then we ask the question. After the search for arguments we set desired proof standards $DV$ for $A(X)$ and $SE$ for $B(X)$ (this cannot be done in advance, as before the search those statements don't exist in any argumentation graph). Response of *Carneades*'s abductive construction of arguments, shown in figure 2, presents the full[3] information regarding this issue. The statement happens to be acceptable, because there is no defensible con argument and there is a pro argument from $B(X)$, which in turn is acceptable, because its proof standard ($SE$) is satisfied even in the presense of an argument con $B(X)$.

[3] Given avaliable search resources.

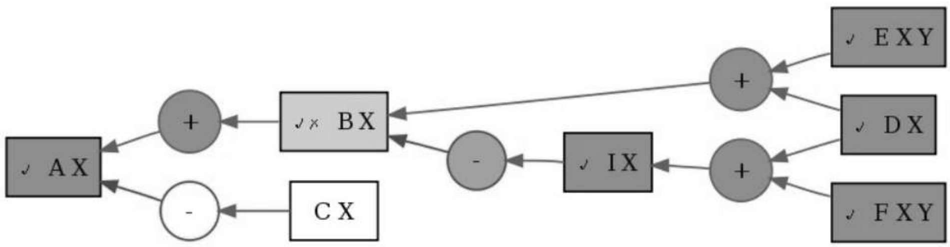



**Figure 2.** ***Carneades*'s Argumentation graph for $A(X)$ with proof standard $DV$**

Now, let us consider the same example using the implementation of $RPA*$. The above knowledge base, written down in the textual format described in section 5, looks as follows:

```
r1: B(x) -> A(x)
r2: D(x), E(x, y) -> B(x)
r3: I(x) -< B(x)
r4: D(x), F(x, y) -> I(x)
r5: C(x) -< A(x)
r6: G(x), E(x, y) -> C(x)
B(X) p=SE
D(X) s=ACCEPTED
E(X, Y) s=ACCEPTED
F(X, Y) s=ACCEPTED
```

The information about proof standards is given in advance. After asking the query $A(X)$ with proof standard $DV$ the application returns result shown in figure 3, which is the minimal information required to determine acceptability of $A(X)$.

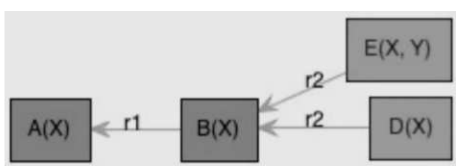



**Figure 3. Result for query $A(X)$ with proof standard $DV$ returned by $RPA*$**

When we deal with relatively small cases, it is best for argumentation analysis to have as much information as possible avaliable at once. However, in argumentation cases more close to real-life situations, both the size of resulting graph and the time spent on it's computation become significant enough to consider incremental analysis of the case using partial information provided e.g. by the algorithm presented here.

### Conclusions and future work

This article presents an algorithm for incremental argumentation analysis in *Carneades*. It is proposed, that such analysis can be conducted using a search algorithm for finding a minimal argumentation graph that allows for determining acceptability of the given goal. Because steps in the search space are considered potentially expensive (e.g. involving OWL reasoner), it is important to complete the search in the minimal number of steps.

The solution for incremental argumentation analysis consists of three steps: (a) proposing a defeasible logic (called RCAF) based on *Carneades*; (b) proposing a general approach for construction of RCAF reasoners by formulating the problem of inference in RCAF as a generic search task; (c) designing and implementing a sample algorithm for reasoning in RCAF, which is $RPA*$ It uses heuristics and search constraints for choosing the exploration paths of argumentation search space.

It is claimed that this algorithm can be helpfull in argumentation analysis in cases where full argumentation graphs are too large to grasp for the application user and in situations where it takes a long time to generate them.

As of the end of 2010, the implementation of the newest version of *Carneades* (as presented in [8] and [2]) is avaliable at the project's website. Future work will mainly focus on integrating the $RPA*$ algorithm with this implementation which will allow its more extended evaluation using resources avaliable in the main *Carneades* project.

### References

[1] Jarosław Arabas and Paweł Cichosz, Search-Based View of Artificial Intelligence. In *Artificial Intelligence Studies*, volume 26, pages 13–29, Siedlce 2006.

[2] S. Ballnat and T. F. Gordon. Goal selection in argumentation processes – a formal model of abduction in argument evaluation structures. In *Computational Models of Argument – Proceedings of COMMA 2010*, pages 51–62, IOS Press, 2010.

[3] Thomas Gordon. Hybrid reasoning with argumentation schemes. In *8th Workshop on Computational Models of Natural Argument*, 2008.

[4] Thomas Gordon. Report on prototype decision support system for oss license compatibility issues. Technical report, Fraunhofer FOKUS, Berlin, 2010.

[5] Thomas F. Gordon and Douglas Walton. The Carneades argumentation framework – using presumptions and exceptions to model critical questions. In Paul E. Dunne and Trevor J. M. Bench-Capon, editors, *Computational Models of Argument – Proceedings of COMMA 2006*, volume 144, of *Frontiers in Artificial Intelligence and Applications*, pages 195–207, Amsterdam, 2006, COMMA, IOS Press.

[6] Thomas G. Gordon and Henry Prakken, and Douglas Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10–15): 875–896, 2007.

[7] Judea Pearl. *Heuristics: intelligent search strategies for computer problem solving*. Addison-Wesley Longman Publishing Co., Inc., 1984.

[8] Iyad Rahwan and Guillermo R. Simari. *Argumentation in Artificial Intelligence*. Springer Publishing Company, Incorporated, 1st edition, 2009.

[9] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, 2003.

Paweł Łoziński
Institute of Computer Science,
Warsaw University of Technology,
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland;
plozinsk@ii.pw.edu.pl

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 23 (36) 2011

**Edward Bryniarski, Zbigniew Bonikowski,**
**Jacek Waldmajer, Urszula Wybraniec-Skardowska**
Opole University, Poland

# REALISTIC PREMISES OF EPISTEMIC ARGUMENTATION FOR DYNAMIC EPISTEMIC LOGICS

**Abstract**: In the paper, certain rational postulates for protocols describing real communicating are introduced. These rational postulates, on the one hand, allow assigning a certain typology of real systems of interactions, which is consistent with the reality of epistemic argumentation in systems of communicating, and on the other one – defining rules of using argumentation in real situations. Moreover, the presented postulates for protocols characterize information networks and administering knowledge in real interactivity systems.
Due to the epistemic character of the considerations, the problem undertaken in the paper concerns working out fundamental assumptions that refer to building of epistemic logics. They allow establishing the correctness of the discourse defined by rational postulates of protocols of real communication. In the context of the presented problem there are the following two research questions distinguished: 1) How do we determine the rule of building of real dynamic epistemic logics? and 2) How should we define semantics for these logics? Within the framework of considerations relating to the research questions asked, certain epistemic operators, relativized to types of communicating, are introduced. Basic logical relations between using these operators are established for these operators. The relations are presented by a diagram called the square of epistemic operators. On the basis of these logical relations some axioms for real dynamic epistemic logics are presented. The semantics of real dynamic epistemic logics is extended by the methods of lower and upper approximation of formula evaluating. This allows defining 'approximation Kripke models'. The results of conceptualization of knowledge on real premises of epistemic argumentation presented in this paper can be applied to rhetoric in real systems of interaction.

**Keywords**: postulates for protocols of epistemic argumentation, epistemic argument and argumentation, system of communicating; basic types of communicating determined by input/output attributes, square of epistemic operators for different aspects of knowledge, approximate semantics, approximation Kripke model, epistemic rhetoric.

## Introduction

Presenting the problem of building of epistemic logics in the context of affecting rhetorical argumentation was inspired by current research of Johan van Benthem [2] and his scientific group, as well as by a certain research

ISBN 978–83–7431–305–6 ISSN 0860-150X 

approach, especially presented by Witold Marciszewski in [9] and expressed by the following utterance (see Preface, p. vii–viii):

> The intended outcome dealt with by rhetoric is the change of certain cognitive state of an addressee effected by a cognitive state of *an addresseer* with the use of a spoken or written text. This definition is enough to show the input to cognitive science to be expended from rhetoric. [...]
> Rhetoric in the version designed in this essay as *cognitive rhetoric* is that theory of communicative interaction whose core involves the issues of rational argument.

In this paper we will give postulates for protocols of *epistemic argumentation* corresponding to real premises of rhetorical argumentation used in epistemic reasoning. Protocols determine rules of using argumentation in real situations. They make up an **epistemic motivation to use argumentation** (they establish **toposes**). They also establish attributes that allow choosing a suitable epistemic logic for using effective argumentation. They are simultaneously a rational means of using knowledge to argumentation with the aim to influence conceptual processes. We will consider epistemic logic in a dynamic approach with regard to dynamic shaping of conceptual systems and vagueness of notions. This approach can be applied by rhetors in order to use transitions and means of a composition of argumentation in an appropriate way.

Determination of administering knowledge by a rational agent acting in compliance with certain protocols of argumentation within a real system of interaction, a system of communicating, requires postulating rationality of acting by the agent, as well as postulating restriction of this rationality appropriately to the real actions.

Generally, a description of administering knowledge by the rational agent in compliance with *dynamic epistemic logics* (*DEL*) protocols was presented in [1], [2], [3], [4], [6], [7], [8].

The notion of bounded rationality was introduced in the 20$^{\text{th}}$ century by H. A. Simon [11], who proposed to distinguish: (1) a set of agents, (2) a set of behaviour alternatives, (3) a set of outcomes of choice among the behaviour alternatives, and (4) a set of order of preferences for making choices of behaviours. According to him, an agent who is invested with "perfect rationality" possesses a full knowledge of distinguished sets, whereas an agent with bounded rationality, in contrast, might not know all alternatives; nor does he need to know the exact outcome of each. What is more, such an agent might lack a complete preference ordering, which is indispensable to obtain the outcomes.

We assume that establishing a proper *DEL* protocol for the agent with bounded rationality leads to linking the real system of interactions with relevant types of communicating. Thanks to fixing the type of communicating, it becomes possible to assign a suitable class of Kripke models for *DEL* to this type. In a real system of interactions, a set of rational agents is limited to a set of subjects of such actions of communicating as: production, rendering available and possession or allocating the objects distinguished by agents. The objects are products of the action of communicating. The products are divided into resources, goods, services and values arising in consequence of actions realized within the real system of interactions. The order of preferences for making choices of actions necessary to obtain certain products expected by agents as a result of a given action, is determined by real conditions that establish the beginning and the end of this action. The indicated context of considerations leads to putting forward the following question: How can we build epistemic logics that allow establishing the correctness of a discourse defined by rational postulates of real communication protocols? Solving the above problem requires, among others, acceptance of a protocol which settles how the rules of building real dynamic epistemic logics and approximated semantics for these logics should be determined.

In this paper, we will present rational postulates which allow executing a certain typology of real systems of interactions. They are divided into the following four groups:

- Postulates for protocols concernig information networks (***P0–P3***),
- Postulates for protocols of the real interactivity system (***P4–P8***)
- Postulates for protocols of administering knowledge (***P9–P11***),
- A postulate for protocols of approximated semantics for *Real-DEL* (***P12***).

These postulates will be introduced in successive sections of the paper.

## 1. Postulates for protocols concerning information networks

***P0. An epistemic argument transfers information about one or many objects in interactive communication.*** *During communicating this information results in accepting or rejecting certain information about these objects.*

***P1. Information about an object O*** *(in short:* ***information****) is a sequent of data about the object* ***O****, or more precisely – a sequent of data*

***identifying*** *the object* ***O*** *or any object being part of the object* ***O****. Pieces of information are* ***indiscernible*** *when they identify the same objects. Identification of an object* ***O*** *groups information about the object* ***O****, thus it groups indiscernible pieces of information.*

***P2.*** *Epistemic arguments refer to a connection of information about objects. Such* ***reference of information about objects*** *are tuples of information about objects. The first piece of information in the given reference identifies the object which the last piece of information is about in this reference.*

***P3.*** *Epistemic argumentation is an intended transmission and processing of information. References on elements determining the same object* ***transmit information*** *on this object. The first element of this reference is a piece of* ***input*** *information, while the last one –* ***output*** *information. References not only transmit information, but also* ***process information****: the first piece of information – the* ***input*** *one – into the last piece of reference information – the* ***output*** *one. Information transmission is a particular case of information processing. We call the object which assigns ordered systems of objects to references an* ***information channel****. The first object of the system determined by the information channel is the* ***input of the channel****, while the last object of this system – the* ***output of the channel****. The information channel processes information if each n-th piece of information of reference determines the n-th object of the system of objects ordered by this channel system of objects. We call the collection of information channels an* ***information network****. The inputs and outputs of information channels will be called the inputs and outputs of the information network. The Internet is a model example of an information network. An information network is also recognized in a discourse, in particular, in a dialogue or a discussion.*

## 2. Postulates for protocols of the real interactivity system

Any language communication is held within a real interactivity system. We will understand the **real interactivity system** as a system of communicating, whose model example is the Internet. In such a system, processing information means producing **resources** of knowledge and respective rendering them available, which leads to possessing or allocating of the knowledge, for instance, producing, rendering available, possessing or allocating of files which include some data or serve the purpose of processing these data. Production and making available of the resources of knowledge, according to common needs of users of the system, is a certain **good** provided for the users by informatics. Production and rendering available of the resources of knowledge, as requested by the users in order to satisfy individual needs, is – for the users – a certain **service** provided by informatics. The equivalent usefulness of resources, goods and services establishes their **value** for users of the communication system. Possession or allocation of accessibility to the goods and services, as well as to the value is – at the same time – a process of producing new information resources.

***P4*. *Argumentation occurs in a system of communicating.*** *A* ***system of communicating*** *is a system of human activity and – at the same time – an information network defined for sets of objects that are subjects or objects of production, rendering available and possession or allocation of resources, goods, services and values being effects of people's informatics-related activity within the system. Still, each input and output of this information network is a subject of production, rendering available, possession or allocation.* ***Knowledge*** *is information processed in a certain system of communicating. A set of data on the subject, relating to the kind of knowledge that the subject possesses, is understood to be information about the subject.* ***Communicating*** *is processing information within the system of communicating. Pairs of such attributes of input/output, subjects' activity at the inputs and outputs of the communicating system as production, rendering available, possession or allocation allow distinguishing the basic types of communicating. We accept that the informatics-related activity of those communicating with one another, which is determined by the above-listed attributes points – with the dominance of this activity – to only one type of their activity. We accept that communicating is as follows:*

***Interactive*** *(with index 1) – when, at the input, there dominates production of knowledge of the net user, while – at the output – this knowledge is rendered available to the user, e.g. ordering to have money transferred to the bank account, in consequence of which the knowledge about the operation made is made available on the account, or the other way round: when at the input one net user renders available knowledge to another user at the output in order to process it, e.g. logging on the bank account and calculating – with the use of the calculator accessible there – the interest rate on the credits granted,*

***Verbal*** *(with index 2) – when, at the input, there dominates possession of knowledge, while – at the output – allocation of the knowledge, or the other way round – one of the users possesses knowledge (e.g. on a website) of another user, or the other way round – on the website of the first net user there is allocated knowledge which the other user possesses in his computer, this knowledge is automatically acquired from the computer of the other user; let us note that this kind of communicating can occur without referring to the meaning of sentences which represent the processed knowledge (content of the information), therefore this communicating can be called verbal,*

***Public*** *(with index 3) – when at the input and at the output there dominates allocation of knowledge, e.g. readers of a published title, by means of questionnaires meant to examine what kind of knowledge they allocate, cause the editors of the title – after getting acquainted with the questionnaires – to allocate and present this knowledge in the title they edit; it also happens that titles – through presentation of the allocated knowledge – influence the type of knowledge their readers will allocate,*

***Private*** *(with index 4) – when at the input and at the output there dominates possession of knowledge, which most often takes place while transferring personal data, e.g. the data are passed when the provider of a service must possess the data which the receiver of the service does; in a similar way a person's identity card is displayed to a police officer,*

***Static*** *(with index 5) – when at the input there dominates rendering knowledge available and at the output – allocation of knowledge, e.g. an Internet website displays a road map and the Internet user – on the basis of the map – allocates knowledge about roads to reach Copenhagen; or the other way round – when at the input there dominates allocation of knowledge, while at the output – rendering it available, e.g. the Internet user renders knowledge allocated by an Internet forum on the very forum itself; in the process of communicating no new data are produced (the data are only made available and are allocated),*

***Dynamic*** *(with index 6) – when at the input there dominates production of knowledge, while at the output – possession of knowledge, e.g. one of the communicating subjects produces new data in order to change the resources of knowledge of the other subject; or the other way round – at the input there dominates possession of knowledge, while at the output – production of knowledge, e.g. the subject, at the output, makes use of knowledge of the other subject in order to make alterations,*

***Decision-making*** *(with index 7) – when at the input and at the output there dominates production of knowledge – the first subject of communication changes the data in the way such that the other of the subjects could implement the changes to make his own alterations; it can also be otherwise – the other subject will be able to make appropriate alterations of data obtained in the process of communicating then and only then when the first subject makes relevant changes of the data; thus, the changes being made depend on decisions on making the changes undertaken by the subjects,*

***Discursive*** *(with index 8) – when at the input and at the output there dominates rendering knowledge available, which most often takes place in a discourse, i.e. when two subjects communicating with each other process knowledge in order to mutually make it available,*

***Intelligent*** *(with index 9) – when at the input there dominates production of knowledge, while at the output – allocation of knowledge and such production of knowledge that by the first subject that the knowledge could be allocated by the other subject,*
*or the other way round – when at the input there dominates such allocation of knowledge by the first subject that the other subject could produce something out of it at the output; both of the described actions can be considered a manifestation of intelligence,*

***Behavioural*** *(with index 10) – when at the input there dominates rendering knowledge available and at the output – possession of knowledge, e.g. if the first of the subjects holds a lower social rank than the other subject (is dependent on the other one), then the first of the subjects must make knowledge available to the other in order that the latter would expand his knowledge,*
*or the other way round – when at the input there dominates possession of knowledge, while at the output – rendering knowledge available, e.g. if the first subject has a higher social rank than the other (the other subject is dependent on the first), then the first subject must possess knowledge which can be rendered available to the other one in order that the rank of the former could be established.*

*We accept that the above-mentioned types of communicating are disjoint in the aspect of subjects' activity: if, between two subjects, there occurs communicating of one of the types, then the other types of communicating do not occur.*

***P5.*** ***Epistemic agent*** *(in short:* ***agent****) is an object at the input or output of a system of communicating.*

**Table 1. Types of communicating determined by input/output attributes**

| input\output | Production | Rendering avail. | Possession | Allocation |
|---|---|---|---|---|
| Production | Decision-making 7 | Interactive 1 | Dynamic 6 | Intelligent 9 |
| Rendering avail. | Interactive 1 | Discursive 8 | Behavioural 10 | Static 5 |
| Possession | Dynamic 6 | Behavioural 10 | Private 4 | Verbal 2 |
| Allocation | Intelligent 9 | Static 5 | Verbal 2 | Public 3 |

*The opposition of the types is represented by means of the following juxtapositions of textures of opposing patterns (opposing colours): (9, 10), (4, 3), (1, 2), (6, 5), (7, 8).*

***P6***. *The following* ***aspects of knowledge*** *are distinguished*:

***Common-sense knowledge*** – applied knowledge and habitual knowledge, which – for agent $a$ is distinguished by the operator of assertiveness ($\boldsymbol{A_a}$): ***agent*** $a$ ***thinks that ....***

***Emotive knowledge*** – knowledge related to feelings distinguished for agent $a$ by the operator of feeling ($\boldsymbol{F_a}$): ***agent*** $a$ ***feels that ....***

***Sensual knowledge*** – knowledge obtained through perception, not experienced or verified, creating an image of objects perceived, distinguished for agent $a$ by the operator of perception ($\boldsymbol{P_a}$): ***agent*** $a$ ***perceives that ....***

***Empirical knowledge*** – not a sensual type of knowledge, yet knowledge attained through experiencing, verifying, testing components of sensual knowledge, distinguished for agent $a$ by the operator of experience ($\boldsymbol{E_a}$): ***agent*** $a$ ***experiences that ....***

***Rational knowledge*** – knowledge attained through thinking and reasoning distinguished for agent $a$ by the operator of understanding ($\boldsymbol{K_a}$): ***agent*** $a$ ***knows that ....***

***The rational knowledge consists of the above-listed aspects of knowledge, as well as types of knowledge defined through relations between the above aspects of knowledge***:

*I know that $\varphi$ if*

(***alternative of the aspects of knowledge***)

*I think that $\varphi$ or my feeling is that $\varphi$ or I perceive that $\varphi$ or I experience that $\varphi$;*

(***principle of subordination***)

*when I think on the basis of experience or feel on the basis of perceiving;*

(***principle of oppositions***)

*if I think, I do not feel,*
*if I feel, I do not think,*
*if I experience, then I do not perceive,*
*if I perceive, then I do not experience;*

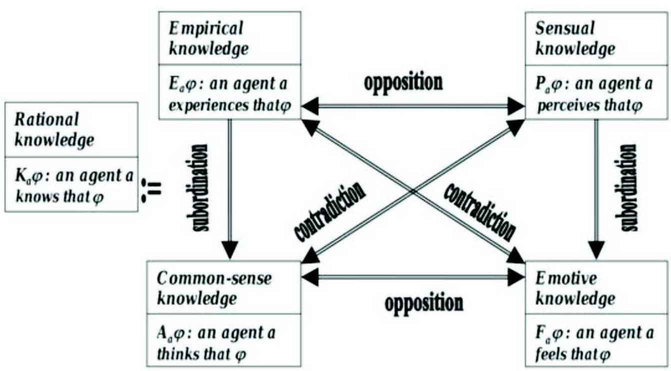
(***principle of contradiction***)

*I do not think iff I perceive, I do not experience iff I feel.*

The above-mentioned aspects of knowledge and types of communicating, defined earlier, allow us to communicate and to define bounded activities of agents in practice. These restrictions can be established, making relevant observation of agents communicating and using such suitable research methods as making polls, testing, computer simulation and so on. The results of this research also offer a reliable image of agents' interaction, leading to showing the real system of interaction.

The basic epistemic operators applied in the real system of interaction satisfy the following logical square given in Diagram 1:

**Diagram 1. Square of epistemic operators for different aspects of knowledge**




Juxtaposing the fundamental epistemic operators with types of communication with indexes 1–10, we obtain the following matrix of epistemic operators:

**Matrix of epistemic operators**

| type/aspect | 1. $A_a$ | 2. $F_a$ | 3. $P_a$ | 4. $E_a$ | 5. $K_a$ |
|---|---|---|---|---|---|
| 1. Interactive | $A_a^1$ | $F_a^1$ | $P_a^1$ | $E_a^1$ | $K_a^1$ |
| 2. Verbal | $A_a^2$ | $F_a^2$ | $P_a^2$ | $E_a^2$ | $K_a^2$ |
| 3. Public | $A_a^3$ | $F_a^3$ | $P_a^3$ | $E_a^3$ | $K_a^3$ |
| 4. Private | $A_a^4$ | $F_a^4$ | $P_a^4$ | $E_a^4$ | $K_a^4$ |
| 5. Statistic | $A_a^5$ | $F_a^5$ | $P_a^5$ | $E_a^5$ | $K_a^5$ |
| 6. Dynamic | $A_a^6$ | $F_a^6$ | $P_a^6$ | $E_a^6$ | $K_a^6$ |
| 7. Decision-mak. | $A_a^7$ | $F_a^7$ | $P_a^7$ | $E_a^7$ | $K_a^7$ |
| 8. Discursive | $A_a^8$ | $F_a^8$ | $P_a^8$ | $E_a^8$ | $K_a^8$ |
| 9. Inteligent | $A_a^9$ | $F_a^9$ | $P_a^9$ | $E_a^9$ | $K_a^9$ |
| 10. Behavioral | $A_a^{10}$ | $F_a^{10}$ | $P_a^{10}$ | $E_a^{10}$ | $K_a^{10}$ |

For any set of agents, set of types of communicating and sets of aspects of knowledge processed in the communicating process there exists relevant *DEL* with epistemic operators determined by types of communicating and aspects of knowledge (as in the matrix of epistemic operators). These logics can be called *Real-DEL*.

**Proposed axioms for *Real–DEL*:**

**Subordination**

$E_a^i\varphi \Rightarrow A_a^i\varphi$

$P_a^i\varphi \Rightarrow F_a^i\varphi$

**Contradiction**

$E_a^i\varphi \Leftrightarrow \neg F_a^i\varphi$

$A_a^i\varphi \Leftrightarrow \neg P_a^i\varphi$

**Oposition**

$E_a^i\varphi \Rightarrow \neg P_a^i\varphi$

$P_a^i\varphi \Rightarrow \neg E_a^i\varphi$

$A_a^i\varphi \Rightarrow \neg F_a^i\varphi$

$F_a^i\varphi \Rightarrow \neg A_a^i\varphi$

**Alternative of the aspects of knowledge**

$A_a^i\varphi \vee P_a^i\varphi \vee E_a^i\varphi \vee F_a^i\varphi \Rightarrow K_a^i\varphi$

***P7*. *Administering knowledge*** *is processing knowledge within information channels in which communicating occurs. It follows from the definition of the information channel and determining the agent that the input and the output of the information channel is a certain agent. Information channels which compose administering the knowledge are* ***dispositions of knowledge****. The fact that the agent knows some-*

*thing, encodes, decodes and represents knowledge, acquires knowledge, announces knowledge, is convinced (believes in something), is interpreted as making use of suitable dispositions of knowledge by the agent: possessing knowledge, encoding, decoding, etc. We call the whole of administering the knowledge the **state of administering knowledge** (in short: **state**).*

***P8***. *In order to administer knowledge, a group of agents who realize a certain type of communicating accept an appropriate **protocol of processing knowledge** that implements this type of communicating.*

## 3. Administering resources of knowledge

The presented rational postulates for *DEL* allow establishing sets $S$ of all states of administering knowledge within the selected real system of interaction. Let $\boldsymbol{P}$ be a set of atomic propositions expressing knowledge, and $\boldsymbol{A}$ be a set of agents. Relations of using – by agents – information channels, are then determined by the mapping $R_{\boldsymbol{A}} : \boldsymbol{A} \rightarrow \wp(S \times S)$, and also the mapping $V^{\boldsymbol{P}} : \boldsymbol{P} \rightarrow \wp(S)$ is known as it determines a set of states, in which for the given atomic proposition there occurs communicating that processes this atomic proposition. Structure $M = \langle S, R_{\boldsymbol{A}}, V^{\boldsymbol{P}} \rangle$ is then a *Kripke model for DEL* (cf. [6]).

Let us note that determining the real system of interaction is executed in a certain relational data basis. The above-mentioned postulates allow identifying attributes of this data basis and values of these attributes. This aspect of the research offers the possibility, in the case of vagueness in determining results of communicating, of applying the method of rough sets in Pawlak's sense [10] to describe this communicating. Administering resources of knowledge in social and economic systems of managing knowledge can be described in this sense as relational data bases, and then – by means of these bases – certain classes of Kripke models can be fixed for *DEL*. A result of such research can be fixing of this type of *DEL* for the given system of managing knowledge. The rational actions proposed here, which lead to fixing certain classes of models for *DEL*, can be made precise by accepting the following postulates:

***P9***. *Protocols of processing knowledge must be established for each type of communicating so that the agents communicating (within this type) could administer, in certain states, a set of atomic sentences that are true only within this type of communicating: with the established* *semantics of DEL other sentences can also be processed within this type of communicating and acknowledged or not to be true.*

***P10***. *The set $S$ of states of administering knowledge is a sum of disjoint sets $S_1, S_2, \ldots, S_{10}$, and $S_i$ – in compliance with* ***P9*** *– corresponds to the type of communication with index $i$ given in* ***P4***.

***P11***. *In the language of DEL there are distinguished epistemic operators: assertive $A_a^i$, of feeling $F_a^i$, perception $P_a^i$, experiencing $E_a^i$, understanding $K_a^i$, where each operator, respectively (as in* ***P9****), distinguishes atomic sentences in the i-th type of communicating.*

## 4. Approximate semantics for *Real–DEL*

The truthfulness of the formula of *DEL* language in model $M = \langle S, R_{\boldsymbol{A}}, V^{\boldsymbol{P}} \rangle$ can be defined in an equivalent way to the standard definition through an extension of valuation function $V^{\boldsymbol{P}} : \boldsymbol{P} \to \wp(S)$ to function $V : \boldsymbol{FORM} \to \wp(S)$, where $\boldsymbol{FORM}$ is a set of properly built *DEL* formulas so that for any formula $\varphi \in \boldsymbol{FORM}$

$$M, s \models \varphi \text{ iff } s \in V(\varphi).$$

Accepting postulates ***P9*** and ***P10*** one can ask the question in what way sets $V(\varphi)$ of *states of administering knowledge* depend on sets $S_1, S_2, \ldots, S_{10}$, that is what the relationship between types of communicating and truthfulness of formulas is. An answer to this question can be obtained by using the method of rough sets in Pawlak's sense [10]:

***P12***. *Assessing set $X = V(\varphi)$* ***from the bottom (as a lower approximation)*** *by means of the set*

$$A^-(X) = \bigcup\{S_i : S_i \subseteq X, \quad i = 1, 2, \ldots, 10\},$$

*and also* ***from the top*** *(as an upper approximation) by means of the set*

$$A^+(X) = \bigcup\{S_i : S_i \cap X \neq \emptyset, \quad i = 1, 2, \ldots, 10\}$$

*we can determine the relation of types of communicating and truthfulness of formulas in the following way:*

*Truthfulness of the two formulas $\varphi, \psi$, does not depend on a choice of type of communication, when*

$$A^-(V(\varphi)) = A^-(V(\psi)),$$
$$A^+(V(\varphi)) = A^+(V(\psi)).$$

Logical values of formulas $\varphi$ and $\psi$ are *indiscernible* (equivalent) in all types of communicating, symbolically: $V(\varphi) \approx V(\psi)$ if their lower approximations and their upper approximations are the same.

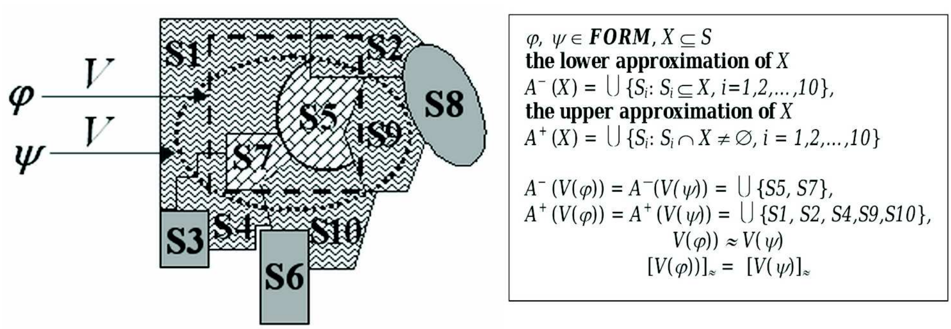
Postulate ***P12*** allows existence of equivalence classes $[V(\varphi)]_{\approx}$ of sets $X \subseteq S$ of states of administrating knowledge such that

$$A^{-}(V(\varphi)) = A^{-}(X),$$
$$A^{+}(V(\varphi)) = A^{+}(X)$$

for any formula $\varphi \in \boldsymbol{FORM}$.

The following Diagram 2 illustrates the above-given method of approximation of logical values of formulas $\varphi$ and $\psi$.

**Diagram 2.**




The square and the ellipse represent sets of states: the logical values of formulas $\varphi$ and $\psi$, respectively; the wavy part of the diagram corresponds to the upper approximation, while the checked part – to the lower approximation of these values. Inside the box on the right, there are definitions of the lower approximation and the upper approximation of sets of states; an equivalence relation $\approx$ defined on sets of states is also determined. The last equation expresses the identity of equivalence classes for equivalent sets of states (values of logical formulas $\varphi$ and $\psi$. In 1982, Zdzisław Pawlak called equivalence classes defined in an analogous manner – rough sets.

Accepting postulates ***P1–P12***, the mapping defined by the following formula:

$$[V] : \boldsymbol{FORM} \rightarrow \{[V(\varphi)]_{\approx} : \varphi \in \boldsymbol{FORM}\},$$

can be called the ***approximation valuation***, and the structure $[M] = \langle S, R_{\boldsymbol{A}}, V^{\boldsymbol{P}}, [V] \rangle$ can be called the ***approximation Kripke model***.

**Towards epistemic rhetoric**

The results of conceptualization of knowledge on real premises of epistemic argumentation presented in this paper can be applied precisely to rhetoric in real systems of interaction. The indicated method of building different types of Kripke's models for dynamic epistemic logics can also be applied to building different models for persuasive aspects of argumentation (see [5]). This is a way leading to "epistemic rhetoric" serving to influence epistemic reasoning.

## References

[1] J. van Benthem: Merging Observation and Access in Dynamic Epistemic Logic, *Studies in Logic* 1(1) (2008), 1–16.

[2] J. van Benthem: *Logical Dynamics of Information and Interaction*, Cambridge University Press, Cambridge 2010.

[3] J. van Benthem, J. Gerbrandy, T. Hoshi, E. Pacuit: Merging Frameworks for Interaction, *Journal of Philosophical Logic*, 38(5) (2009), 491–526.

[4] J. van Benthem, M. Martinez: The Logical Stories of Information, in: J. van Benthem and P. Adriaans (eds.) *Handbook of the Philosophy of Information*, Elsevier Science Publishers, Amsterdam 2008, 217–270.

[5] K. Budzyńska, M. Kacprzak: Formal Models for Persuasive Aspects of Argumentation, *Studies in Logic, Grammar and Rhetoric* 16 (29) 2009, 159–187.

[6] H. van Ditmarsch, W. van der Hoek, B. Kooi: *Dynamic Epistemic Logic.* Synthese Library 337, Springer, Dordrecht 2008.

[7] X. He, J. Horty, E. Pacuit (eds.): *Logic, Rationality and Interaction*, Proceedings of the Second International Workshop (LORI), Chongqing, China, October 2009, Springer-Verlag, Berlin-Heidelberg 2009.

[8] T. Hoshi: *Epistemic Dynamics and Protocol Information*, Ph.D. thesis, Department of Philosophy, Stanford University, (ILLC-DS-2009-08).

[9] W. Marciszewski, *Logic from a Rhetorical Points of View*, Walter de Gruyter, Berlin–New York 1994.

[10] Z. Pawlak: *Rough Sets. Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publishers, Dordrecht 1991.

[11] H. A. Simon: A Behavioral Model of Rational Choice, *Quarterly Journal of Economics* 69 (1955), 99–118.

Edward Bryniarski
Group of Logic, Language and Information
Institute of Mathematics and Informatics
Opole University
ul. Oleska 48, 45-052 Opole, Poland
edlog@uni.opole.pl

Zbigniew Bonikowski
Group of Logic, Language and Information
Institute of Mathematics and Informatics
Opole University
ul. Oleska 48, 45-052 Opole, Poland
zbonik@math.uni.opole.pl

Jacek Waldmajer
Group of Logic, Language and Information
Institute of Philosophy
Opole University
ul. Katowicka 89, 45-061 Opole, Poland
jwaldmajer@uni.opole.pl

Urszula Wybraniec-Skardowska
Group of Logic, Language and Information
Opole University
ul. Matejki 5/21, 45-055 Opole, Poland
uws@uni.opole.pl

**Jacky Visser**
University of Amsterdam & International Learned Institute for Argumentation Studies

**Floris Bex**
University of Dundee

**Chris Reed**
University of Dundee

**Bart Garssen**
University of Amsterdam & International Learned Institute for Argumentation Studies

# CORRESPONDENCE BETWEEN THE PRAGMA-DIALECTICAL DISCUSSION MODEL AND THE ARGUMENT INTERCHANGE FORMAT

**Abstract**: The pragma-dialectical ideal model of a critical discussion takes a normative approach to argumentative discourse. The model defines the four stages of a critical discussion, conditions on speech acts and their distribution over the stages, and a set of 15 procedural rules regimenting the moves discussants may make. These problem-valid rules are instrumental towards the reasonable resolution of the difference of opinion. We take the model of a critical discussion as constituting a basis for a dialogue protocol allowing agents to play out a dialectical game in order to test the tenability of one agent's standpoint. The Argument Interchange Format (AIF) allows such a dialogue protocol to be translated in terms of its core ontology. The core ontology provides a directed graph data structure in which descriptions of argumentative discourse and arguments can be represented. The AIF can function as interlingua allowing various frameworks and theories of argumentation to interact in theoretically unbiased terms. Establishing a correspondence between pragma-dialectical notions and the AIF would provide the latter with a normative natural language discussion model. Furthermore viewing the pragma-dialectical theory from a formalised perspective indicates possible areas of concern which need to be addressed before the theory could get involved further in the field emerging on the intersection between argumentation theory and artificial intelligence.

**Keywords**: Argument Interchange Format, critical discussion, dialogue protocols, Pragma-Dialectics

## 1. Argumentation and theory

In the last forty years the pragma-dialectical approach to argumentative discourse has been developed into a full-blown argumentation theory and normative discussion model. (van Eemeren and Grootendorst 1984; 2004) The theory takes any argumentative exchange as an instantiation of the

ISBN 978–83–7431–305–6 ISSN 0860-150X 

ideal model of a critical discussion. This allows the discourse to be analysed, reconstructed and evaluated with respect to a normative model. Starting out as a theory based on speech acts as the functional building blocks of linguistic communicative activity ("pragma", short for pragmatics, being the field within linguistics in which meaning is regarded as inherently context-dependent) and a procedure for reasonably resolving a difference of opinion (taking the "dialectical" perspective), it has since been extended to also incorporate rhetorical aims of effectiveness and institutional contexts among others. (van Eemeren 2010) Lately the conventional validity – whether the restrictions in the normative model match accepted conventions in actual use – of the theory has also been put to the test in a series of empirical studies. (van Eemeren, Garssen and Meuffels 2009)

In the past few decades, AI has developed its own sub-field devoted to computational argumentation theory, in which significant theoretical and practical advances are being made. This fecundity, unfortunately, has a negative consequence: with many researchers focusing on different aspects of argumentation, it is increasingly difficult to reintegrate results into a coherent whole. To tackle this problem, the AI community has initiated an effort aimed at building a common ontology for computational argument, which will support interchange between research projects and applications in the area: the Argument Interchange Format (AIF). (Chesñevar *et al.* 2007)

Thus far there has been notably little interaction between computational argumentation theory and the pragma-dialectical approach. In the present paper we will focus on this disciplinary intersection by presenting a preliminary account of the correspondence between the standard pragma-dialectical model of a critical discussion and notions within the AIF.[1] The rules for a critical discussion within the context of the ideal pragma-dialectical discussion model can be taken as constituting the foundations for a dialogue protocol. A justification for the possibility of '*protocolisation*' of the rules can be found in their instrumentality towards the goal of the discussion – i.e. reasonably resolving the difference of opinion. Any move in violation of the rules would obstruct the resolution and would therefore be fallacious. By following such a protocol agents can play a dialectical game in which they decide on the acceptability of a certain proposition in a reasonable manner.

Developing the protocol gives us the opportunity to further investigate the rules for critical discussion on the coherence and consistency of the pro-

[1] The *standard* pragma-dialectical model refers to the theory before its rhetorical extension in terms of strategic manoeuvring. See (van Eemeren and Grootendorst 2004) for the standard model and (van Eemeren 2010) for the extended.

cedure proposed. As such we can investigate the problem-validity of the rules by testing whether all of the rules are actually aimed at the goal of resolving the difference of opinion and whether there are no additional rules necessary to ideally avoid moves that distract from reaching the overall goal.[2] Because of the AIF's links to more formal systems, translating the protocol into the language of the AIF opens up the possibility of actually implementing the dialectical game of a critical discussion in established computational applications and algorithms at a later moment. These can range from tools to visualise argumentation to automated decision-making systems, and from other dialogue games to logical systems that decide on the validity of arguments. From a computational point of view taking pragma-dialectical insights into account can provide a normative foundation to some applications and answer questions such as those raised by McBurney and Parsons (2009) about the design and assessment of dialogue protocols:

> "How many locutions should there be? What types of locutions should be included, e.g., assertions, questions, etc? What are the appropriate rules for the combination of locutions? When should behavior be forbidden, e.g., repeated utterance of one locution? Under what conditions should dialogues be made to terminate?" (p. 275)

Being a normative discussion model the pragma-dialectical theory provides a procedure which regiments moves in deliberative or persuasive dialogues in multi-agent systems. It also presents us with a fully developed overview of admissible locutions and argumentative moves, a speech act based approach that allows for complex, composite speech acts, a notion of discussion stages, of fallacious moves, etc.

The current paper investigates the groundwork of an addition of the pragma-dialectical theory of argumentative discourse to the AIF arsenal as a natural language discussion module. For now we start with a very basic instantiation, creating the opportunity to expand on it in the future. Besides possibly simplifying the theory at points (by, for example, only focussing on single non-mixed differences of opinion – more on which later), we currently steer clear of the rhetorical extension with strategic manoeuvring, the institutional embedding with argumentative activity types and the analysis of argumentative discourse through the use of linguistic indicators and dialectical profiles. (See respectively van Eemeren 2010, and van Eemeren

[2] This is not to say that any problems found would actually be problems to the theory because the specific issue might be addressed in another part of the theory. It could point us towards aspects of the rules that are less well-developed from a formal perspective.

*et al.* 2007) The notion of dialectical profiles interestingly enough appears to be closely linked to what we present in this paper if we regard a dialectical profile or route within the discussion as an instantiation of the possible moves outlined in a critical discussion dialogue protocol and the flow-chart in which our present example has been presented (see Figure 5.) A continuation of the study should take note of these facets of the pragma-dialectical theory and refine the crude correspondences arrived at in what follows. We will first introduce the most relevant aspects of the pragma-dialectical theory and of the AIF in paragraphs 2 and 3. Then we will present a preliminary correspondence between the two in paragraph 4. Paragraph 5 will conclude this paper with an outline of our endeavours so far and of the opportunities it opens up for future research.

## 2. The pragma-dialectical approach to argumentation

### 2.1. The ideal model of a critical discussion

In the pragma-dialectical approach argumentative discourse is analysed relative to the ideal model of a critical discussion. This fully developed discussion model is: *normative*, as opposed to an empirically distinguished dialogue type; takes into account all *stages* of a discussion instead of merely the inference-drawing stage; and pertains primarily to *natural language* discourse in contrast to just arguments expressed in an artificial language devoid of a normative basis for their relation to actual discourse.

According to the pragma-dialectical ideal of reasonableness a critical discussion is aimed at resolving the difference of opinion based on the merits of the respective points of view. In the discussion the parties take on the roles of protagonist and antagonist, respectively arguing for the standpoint or criticising its tenability. Thus they engage in a social interaction aimed at achieving mutual agreement about the (un)acceptability of the proposition expressed in the standpoint.[3] To this avail the discussants perform speech acts and pass through the four stages of a discussion all systematically fulfilling a necessary function in the process of reasonably resolving the difference of opinion. The discussants start off from a set of externalised material and procedural points of agreement, indicating what common ground there is. The dialectical rules ensure a methodical resolution-oriented

[3] Internal deliberation or monologue on this take would be reconstructed as a dialectical process in which both discussion parties are fulfilled by the same individual anticipating on counter moves.

discussion procedure based on these conceded premises – *ex concessis* – by prescribing dialectical obligations and rights to the discussants. The sections that follow will explain the stages (2.2), the speech act distribution (2.3) and the 15 rules (2.4) of a critical discussion.

### 2.2. The stages of a critical discussion

Discussion parties can only resolve their difference of opinion in a reasonable manner if they go about in a well-regimented and systematic manner. In the *confrontation stage* the parties recognise their difference of opinion and externalise it. In a single, non-mixed difference of opinion one of the parties will have committed himself to one particular standpoint which the other party disagrees with. This disagreement is expressed by casting doubt on the standpoint. The disagreeing party can also not merely doubt the standpoint but actually hold an opposite point of view. This would result in a mixed difference of opinion where both discussants have the obligation to defend their own standpoint if they are prompted to do so. There can also be disagreement about several separate but related standpoints at the same time. In such case the difference of opinion becomes multiple. For the remainder of this paper we will focus on single, non-mixed differences of opinion as the elementary case from which more elaborate and complex forms could be composed. The discussion parties will in the *opening stage* agree on a set of mutually accepted premises and procedures, and commit themselves to engage in a critical discussion. At this time they also distribute the roles they will each play in the next stage of the discussion. One of the parties will defend the standpoint at issue as protagonist by putting forward argumentation in support of it.[4] The other party will cast doubt on the standpoint and, as antagonist, will critically challenge the argumentation.[5]

Once these mutual commitments have been made, the *argumentation stage* commences. In this stage the protagonist tries to defend the standpoint by arguing for it, i.e. by performing the complex speech act of argumentation in defence of his standpoint. The antagonist in turn can ask for further

[4] In most instances it will be the advancer of the standpoint who takes on the role of protagonist and the doubter who takes on the role of antagonist, but the parties are free to decide otherwise as would suit their particular situation.

[5] In the sections involving the pragma-dialectical theory the term "argumentation" will be used in a rather specific, technical sense in line with Pragma-Dialectical literature and with its natural meaning in most Roman and Germanic languages. It is taken to denote the constellation of arguments advanced in support of (and not including) a standpoint. It also is the term that names the complex speech act covering the assertives performed in discourse in support of the standpoint expressed.

clarification, question the acceptability or justificatory force of the argumentation – as such soliciting further defence by the protagonist – or he can accept (part of) the protagonist's argumentation. Finally the discussion will enter the *concluding stage* where the current difference of opinion gets resolved by either a retraction of the initial standpoint due to the protagonist's inability to conclusively defend it, or the mutual acceptance of the standpoint due to a defence that was conclusive. Of course if the protagonist has to retract his standpoint this does not mean that the contradiction of the propositional content of it has been constructively argued for.[6] Such would take another critical discussion.

### 2.3. The distribution of speech acts in a critical discussion

The discussants go through the stages of the discussion by performing speech acts. The model of a critical discussion specifies which types of speech acts have to or may be performed by each party at each stage. In analysis, the speech acts that are geared towards the resolution of the difference of opinion constitute the argumentatively relevant utterances that need to be reconstructed. (van Eemeren *et al.* 1993) Assertives are performed to express the initial standpoint and to compose the complex speech act of argumentation in defence of the standpoint. Such a complex speech act is made up of the individual assertions and is at a textual level intrinsically connected to the assertion by which the contested standpoint is advanced. Through commissives the parties accept standpoints and argumentation, and agree on mutual commitments towards common starting points, procedures or the outcome of intersubjective procedures and (sub-)discussions. Directives are used to prompt the other party to defend his standpoint and argue for it. Discussants can always ask for clarification by performing a directive or provide clarification themselves with a usage declarative.[7]

### 2.4. The procedural rules of a critical discussion

The discussion moves discussants may make through performing speech acts while going through the stages of a critical discussion are regimented by 15 rules that ensure a reasonable dialectical procedure. These rules are problem-valid in that obeying them is a necessary condition for reaching the intended outcome of critically testing the standpoint at issue and resolving

[6] Testifying to the *critical rationalist* principles of the theory.

[7] The tables in (van Eemeren *et al.* 2007, p. 16) and (van Eemeren and Grootendorst 1984, p. 105) show the speech acts relevant for critical discussion and their distribution over the discussion stages and between the discussion parties.

the difference of opinion in a reasonable manner. Any violation of the rules for a critical discussion results in a frustration of the resolution procedure and can therefore be called fallacious.[8] We will quickly go through the rules and will reproduce some from (van Eemeren and Grootendorst 2004) if they are of particular interest to our current project.[9] The first of the 15 rules specifies the unconditional right of discussants to advance or cast doubt on any standpoint regarding any proposition regardless of topic or (speaker's) status. The second rule allows the discussant doubting a standpoint to prompt the discussant who advanced the standpoint to actually defend it. Advancing a standpoint in principle commits the discussant to defend it if he is challenged; the burden of proof rests with he who advances a standpoint. There is no such commitment to challenging the standpoint on behalf of the discussant who casted doubt. One provision here is the principle of *non bis in idem*: the proponent of a standpoint is never obligated to defend a particular standpoint if it has already been successfully defended before under the same discussion rules, and premises, against the same opponent. Furthermore a discussion cannot proceed without the discussion parties first agreeing on certain basic rules and premises.

RULE 3:

> *The discussant who is challenged by the other discussant to defend the standpoint that he has put forward in the confrontation stage is always obligated to accept this challenge, unless the other discussant is not prepared to accept certain shared premises and discussion rules; the discussant remains obligated to defend the standpoint as long as he does not retract it and as long as he has not successfully defended it against this particular discussant on the basis of the particular agreed premises and discussion rules.*

During the discussion the parties play the roles of protagonist, defending the standpoint, and antagonist, criticising it. That the discussants need to commit themselves to these roles for the remainder of the current critical discussion is laid out in rule 4. After deciding on the discussion rules, discussants should not digress from them or call them into question again during the current discussion. If a discussant wants to discuss the status of one of

[8] For more on fallacies as violations of the rules of a critical discussion, see (van Eemeren and Grootendorst 1992) and (van Eemeren *et al.* 2002).

[9] The rules as presented here are very similar to those in (van Eemeren and Grootendorst 2004) but are revised slightly in some occasions. Of course the rules of a critical discussion still apply equally to male and female discussants, but in the interest of brevity we use male pronouns to refer to both protagonists and antagonists.

the agreed upon rules this happens outside of the current discussion, giving rise to a meta-discussion.[10]

> RULE 5:
>
> *The discussants who will fulfil the roles of protagonist and antagonist in the argumentation stage agree before the start of the argumentation stage on the rules for the following: how the protagonist is to defend the initial standpoint and how the antagonist is to attack this standpoint, and in which case the protagonist has successfully defended the standpoint and in which case the antagonist has successfully attacked it; the rules in which this is laid down apply throughout the duration of the discussion, and may not be called into question during the discussion itself by either of the parties.*

In the argumentation stage discussants can perform three types of speech acts to critically asses the tenability of the standpoint. First of all the protagonist can perform the complex speech act of argumentation through a constellation of assertives according to rule 6a. This defence of the standpoint is provisional until the antagonist performs a commissive confirming the acceptability of the argumentation. If the antagonist does not accept the argumentation he will perform the illocutionary negation of the commissive and a directive to request new argumentation on the basis of the unacceptability of the propositional content or of the justificatory force of the argumentation to the standpoint (rule 6b).

In case the argumentation is attacked on its propositional content, rule 7 states that the protagonist and antagonist will employ the *intersubjective identification procedure* by checking whether the proposition is part of the set of material starting points which were mutually agreed on in the opening stage. If they agree it is not part of the starting points they can either use a method they specified in the procedural starting points to check the acceptability of the proposition – for example looking it up in an encyclopedia – or they can engage in a sub-discussion with the contested proposition as sub-standpoint.

If the argumentation is attacked on its justificatory (or refutatory) force, rule 8 determines that in the case that the reasoning in the argumentation is fully externalised and is dependent on logical validity, the discussants can proof the validity through the *intersubjective inference procedure* making

[10] Which should not be confused with a sub-discussion. We will encounter the latter in the argumentation stage, while the meta-discussion (also called meta-dialogue by some authors) is used to determine the common commitments of the discussants in the opening stage.

use of the system of logic agreed upon as procedural starting point in the opening stage. Should the argumentation not be dependent on logical validity or fail to be fully externalised it is not logically valid and will make use of an argument scheme. Ordinarily such an argument scheme will not be explicitly stated and will need to be reconstructed. This reconstruction will be carried out by following the *intersubjective explicitisation procedure* which will determine the particular argument scheme employed. Once this has been done, the discussants will have to decide whether the scheme is admissible and has been applied properly. They do this by using the *intersubjective testing procedure*. The admissibility is tested by checking whether this argument scheme and its accompanying critical questions are part of the procedural starting points agreed upon in the opening stage. The application of the scheme is tested by posing the critical questions associated with it and judging whether it can withstand such challenges.

RULE 8:[11]

*a.* *The protagonist has successfully defended a complex speech act of argumentation against an attack by the antagonist with regard to its justificatory (or refutatory) force if the application of the intersubjective inference procedure or (after application of the intersubjective explicitisation procedure) of the intersubjective testing procedure, yields a positive result;*

*b.* *the antagonist has successfully attacked the justificatory (or refutatory) force of a complex speech act of argumentation if the application of the intersubjective inference procedure or (after application of the intersubjective explicitisation procedure) of the intersubjective testing procedure yields a negative result.*

Rule 9 pertains to the conditions of the conclusive attack or defence of a standpoint. The standpoint has been defended conclusively if the antagonist did not manage to successfully attack the propositional content or the justificatory (or refutatory) force of the argumentation in support of this standpoint. The standpoint has been conclusively attacked if the antagonist did manage to successfully attack the content or force of every complex speech acts of argumentation performed by the protagonist in support of this standpoint.

[11] By having a disjunctive form in part *b.* this rule forces the choice we make later in our dialogue protocol when it comes to not regarding argumentation which failed the intersubjective inference procedure as salvageable by employing the intersubjective explicitisation procedure first and then subsequently checking its tenability through the testing procedure.

Although the aim of the critical discussion is to critically test the tenability of a standpoint, the antagonist is under no obligation to attack the argumentation in support of a standpoint in all possible ways. The critical stance of the antagonist can be short-lived if he feels compelled to accept the first attempt the protagonist makes at defending the standpoint. The antagonist does retain the right to critically challenge the argumentation throughout the discussion though as long as he is not repeating himself after a successful defence or an act of retraction with regards to the standpoint or argumentation for it by the protagonist.

Because the protagonist should defend the standpoint, he has to support it by means of advancing argumentation. Quite similar to the antagonist's right expressed in rule 10, the protagonist retains the right to defend his argumentation throughout the discussion. Should an argumentation be attacked on both its propositional content and its justificatory force, then the protagonist has to defend against both. Aside from the right to defend a proposed argumentation against attacks, rule 12 allows the protagonist to retract the commitment to an argumentation he advanced earlier in order to support the standpoint in a different way.

The rules so far allow for the discussants to frustrate the resolution of their difference of opinion by allowing them to repeat performing the same speech acts over and over again. The orderly conduct of a critical discussion is regulated through rule 13 by posing a restriction on the repetition and mixing of speech act performances and by having the discussants take alternating turns.

In order to end the particular instance of a critical discussion, rule 14 states the pre-conditions for the speech acts discussants may perform in the concluding stage of the discussion. The discussants will decide on the outcome of the discussion leading the protagonist to have to retract his standpoint if it has not been conclusively argued for or leading to the antagonist having to retract his doubt regarding the standpoint if it has. Although rule 14 allows for an outcome of the discussion in which none of the discussants has to change their commitment to the standpoint, such a termination can not be regarded an instance of a reasonably resolved difference of opinion.

Because of the nature of the dialectical procedure (i.e. being based on externalised commitments) it is very important that the discussion parties optimally formulate and interpret their utterances. The utterances should further the resolution process, not obstruct it. To this end, discussants may always perform a usage declarative themselves or ask their dialectical opponent to do so, in which case the other is obligated to comply.

This concludes the normative 15 rules of a critical discussion as well as our present introduction of the pragma-dialectical theory. In paragraph 4 we will establish some basic correspondences between the pragma-dialectical theory we have just seen and the Argument Interchange Format which will be introduced in paragraph 3.

## 3. The Argument Interchange Format

Argumentation theory is a large and diverse field stretching from analytical philosophy to communication theory and social psychology. The computational investigation of the space has multiplied that spectrum by a diversity of its own in semantics, logics and inferential systems. One of the problems associated with the diversity and productivity of the field, however, is fragmentation: with many researchers from various backgrounds focusing on different aspects of argumentation, it is increasingly difficult to reintegrate results into a coherent whole. This in turn makes it difficult for new research to build upon old. To tackle this problem, the computational argument community has initiated an effort aimed at building a common ontology for argument which will support interchange between different research projects and applications in the area: the Argument Interchange Format (AIF).

Owing to its roots in computational argumentation, a main aspiration of the AIF is to facilitate data interchange among various tools and methods for argument analysis, manipulation and visualization.[12] Whilst the ideal of a single format might not be feasible in such a diverse field, a common consensus on the standards and technologies employed is desirable. Furthermore, the AIF project aims to develop a commonly agreed-upon *core ontology* that specifies the basic concepts used to express argumentative information and relations. The purpose of this ontology is not to replace other languages for expressing argument but rather to serve as an abstract *interlingua* that acts as the centrepiece to multiple individual languages for argumentation. These argument languages can be, for example, logical languages (e.g. ASPIC's defeasible logic, see Prakken 2010), visual languages (e.g. Araucaria's AML format for diagrams, see Reed and Rowe 2004) or natural language (e.g. as used in the pragma-dialectical approach, see van Eemeren and Grootendorst 2004).

[12] Even though the AIF has a clear computational objective, such tools and methods need not necessarily be implemented as computer programs: a pragma-dialectical analysis, for instance, is a method that is not implemented as a program.

A common abstract ontology for argumentation is interesting from a practical perspective because it drastically reduces the number of translation functions that are needed for the different argumentation languages to engage with each other; only translation functions to the core AIF ontology have to be defined (i.e., $n$ instead of $n^2$ functions for $n$ argumentation languages). In this way, data interchange is facilitated and methods that use different languages can be applied to the same argument resources expressed in the AIF. With the AIF as an interlingua we can, for example, use a diagramming tool such as Araucaria to visualise arguments that were interpreted from a natural language text using pragma-dialectical methods. From a more theoretical perspective a common ontology is interesting because it provides a conceptual anchoring point for the various different argumentation languages.

### 3.1. The AIF ontology

The AIF is constructed as an 'ontology', which in the context of computer science, and knowledge representation in particular, is a way of defining the key concepts of a domain and the relationships between them. In the AIF ontology, arguments and their mutual relations are described by conceiving of them as an *argument graph.* The ontology falls into two natural halves: the Upper Ontology and the Forms Ontology. The Upper Ontology, introduced in (Chesñevar *et al.* 2007), describes the graphical language of different types of nodes and edges with which argument graphs can be built (i.e. the "syntax" for the abstract language of the AIF ontology). The Forms Ontology, introduced by (Rahwan *et al.* 2007), allows for the conceptual definition of the elements of the graphs, that is, it describes the argumentative concepts instantiated by the elements in a graph (i.e. the "semantics" for our abstract language).

The Upper Ontology places at its core a distinction between *information*, such as propositions and sentences, and *schemes*, general patterns of reasoning such as inference or conflict, which are used to relate pieces of information to each other. Accordingly, there are two types of nodes for building argument graphs, information nodes (*I-nodes*) and scheme nodes (*S-nodes*) and I-nodes can only be connected to other I-nodes via S-nodes. That is, there must be a scheme that expresses the rationale behind the relation between I-nodes. In the basic AIF ontology, scheme nodes can be rule application nodes (*RA-nodes*), which denote specific inference relations, conflict application nodes (*CA-nodes*), which denote specific conflict relations, and preference application nodes (*PA-nodes*), which denote specific preference relations.

The Forms Ontology is important in that it contains the argumentative concepts instantiated by the graph. The Forms Ontology is essentially based on *schemes*, general patterns of reasoning, that is, inference schemes, conflict schemes or preference schemes. Informally, inference schemes are rules of inference, conflict schemes are criteria (declarative specifications) defining conflict (which may be logical or non-logical) and preference schemes express (possibly abstract) criteria of preference. These main scheme types can be further classified. For example, inference schemes can be deductive or defeasible. Defeasible inference schemes can be further subdivided into more specific argumentation schemes, such as the schemes for Causal Argument or for Argument from Sign in (Walton *et al.* 2008) or the pragma-dialectical argument schemes based on analogy, sign or cause (see van Eemeren and Grootendorst 1992).[13] There are various ways to represent the schemes in the Forms Ontology. Rahwan *et al.* (2007), for example, define them as graphs of so-called form-nodes (F-nodes) whilst Rahwan *et al.* (2010) define schemes as combinations of classes of statements in Description Logic. In this paper, we will represent individual schemes as a list of features, viz.

| *Scheme name* | **Analogy** | **Modus Ponens** |
|---|---|---|
| *Scheme type* | defeasible inference scheme | deductive inference scheme |
| *Premises* | *A* is true (false) for *C1*<br>*C1* is similar to *C2* | $\varphi$<br>$\varphi \Rightarrow \psi$ |
| *Conclusion* | *A* is true (false) for *C2* | $\psi$ |
| *Presumption* | The similarity between C1 and C2 is relevant to the comparison | none |
| *Exception* | *A* is false (true) for another *C3* similar to *C1* | none |

**Table 1: Two possible inference schemes in the Forms Ontology**

Note that the critical questions for a scheme are implicitly modelled; some of them point to an implicit presumption ('Is the similarity sufficiently relevant?'), others correspond to the exception ('Is there some other *C3* that is also similar to *C1*, but in which *A* is false?') or they may ask after one of the premises ('Is *A* true for *C1*?').

The Forms Ontology and the Upper Ontology are intimately connected because specific applications of schemes (denoted by RA-, CA- and

[13] It is important to note that the AIF ontology does not (and should not) legislate as to which schemes or forms are the correct ones; different schemes are each plausible according to particular theoretical assumptions.

PA-nodes) are instantiations of general (inference-, conflict- and preference-) schemes; in other words, the S-nodes *fulfil* the schemes expressed in the Forms Ontology. As an example of argument graphs that fulfil schemes consider Figure 1, in which two arguments for Plato's (p) mortality are given, one based on Socrates' (s) mortality and the fact that Plato and Socrates are similar (e.g. they are both men) and another based on the fact that Plato is a man (and therefore mortal). Rectangular nodes are I-nodes and ellipses are S-nodes; the concepts from the Forms Ontology that are fulfilled by the nodes (see the two schemes for Analogy and Modus Ponens above) are rendered next to the nodes.

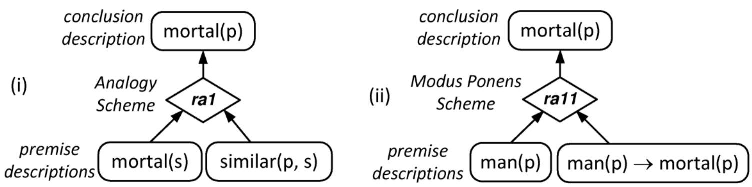



**Figure 1. Argument graphs in the language of the AIF ontology**

### 3.2. Dialogue in the AIF

The basic AIF ontology, as described in (Chesñevar *et al.* 2007; Rahwan *et al.* 2007), does not include ways of representing argument$_2$, that is, dialogical argument.[14] One reason for this is that as Prakken (2005) remarks, while there are a number of well-defined systems for dialogue games, for many of these systems the underlying design principles are mostly implicit. Despite this, Reed *et al.* (2008; 2010) have recently made some tentative steps in the way of including dialogical argument$_2$ in the AIF ontology. The extended ontology, dubbed AIF+, extends the base ontology to support representation of dialogue protocols (i.e. specifications of how dialogues are to proceed), to support representation of dialogue histories (i.e. records of how given dialogues did proceed) and to support representation of the connection between dialogic argument$_2$ and argument$_1$. One underlying premise of this work is that any extensions to the basic AIF should include a minimal amount of extra representational machinery. Below, we briefly summarize the work on the AIF+ ontology.

[14] Here, we refer to O'Keefe's (1977) two characterizations of the term "argument": argument$_1$ and argument$_2$. Argument$_1$ refers to an argument as a static object (the pragma-dialectical notion of *argumentation*) and is described by sentences such as "he prepared an argument". Argument$_2$ refers to a dialogue (the pragma-dialectical notion of *critical discussion*) and is described by sentences such as "they had an argument".

In the context of the AIF+ ontology, it is proposed that locutions are modelled as a subclass of I-nodes called L-nodes. This approach is followed primarily because statements about locution events are propositions that could be used in arguments. So for example, the proposition *Plato says, 'Socrates is mortal'* could be referring to something that happened in a dialogue (and later we shall see how we might therefore wish to reason about its propositional content, *Socrates is mortal*) but it might also play a role in a structure of the form argument$_1$ (say, as a premise in an argument from expert opinion or of an argument about Plato's communicative abilities).

A dialogue is more than a mere sequence of unconnected locutions: there is a functional relationship between different locutions, especially if we consider them in a dialogue with set rules. Imagine, for example, a dialogue in which Plato says, 'Socrates is mortal' and Aristophanes responds by asking, 'Why is that so?' In trying to understand what has happened, one could ask, 'Why did Aristophanes ask his question?' Now, there is at least one answer we could give purely as a result of the dialogue protocol, namely, 'Because Plato had made a statement'. That is to say, there is a functional relationship between the proposition, *Plato says, 'Socrates is mortal'* and the proposition, *Aristophanes asks why it is that Socrates is mortal*. That relationship can be seen as a scheme, a pattern of reasoning (but perhaps not as a conventional inferential scheme as for RA-nodes) of which the grounds lie in the definition of the dialogue game. Thus, by analogy to the ontological machinery of schemes, we can view transitions as Forms that are fulfilled by an S-node for transitions between locutions, which we call transition application nodes (TA-nodes).

Many protocols for dialogue games associate constraints with what are here called transitions. A transition scheme can thus be interpreted as having a *presumption* in much the same way that specific inference schemes have presumptions (cf. the scheme for argument from analogy in Table 1). These transitions and the conditions on them, are not all there is to a protocol: some locutions have conditions which do not directly refer to another locution in the dialogue, that is, constraints on individual locutions. We specify these constraints as pre- and post-conditions on operators that correspond to locutions. Figure 2 shows the ontological structure of locutions and transitions.

For examples of locutions and transition schemes, consider Table 2 and 3, which show the *Challenge* and *Resolve* locutions and the *Challenge-Resolve* transition from Mackenzie's (1979) DC protocol. Notice the difference between constraints-as-presumptions and constraints-as-preconditions: the precondition for a Challenge always holds, no matter to which

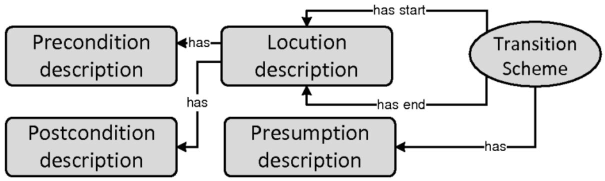



**Figure 2: Transition schemes and locutions**

other locution the Challenge responds. The presumptions on a Challenge-Resolve transition, however, only hold when a Resolve is offered as a response to a Challenge.

| *Locution name* | **Challenge** | **Resolve** |
|---|---|---|
| *Format* | Why P? | Resolve whether P |
| *Precondition description* | P is not in speaker's commitment | none |
| *Postcondition description* | P is in hearer's commitment<br>Why P? is in speaker's commitment | none |

**Table 2: Two locutions from Mackenzie's DC protocol**

| *Scheme name* | **Challenge – Resolve** |
|---|---|
| *Start Locution Description* | Why P? |
| *End Locution Description* | Resolve whether 'if Q then P' |
| *Presumption Description* | P is an immediate consequence of Q<br>Q is a conjunction of statements to all of which the hearer is committed |

**Table 3: A transition in Mackenzie's DC protocol**

One interesting question is how exactly L-nodes are connected to I-nodes in argument$_1$. So, for example, what is the relationship between the proposition *Socrates is mortal* and the proposition *Plato says, 'Socrates is mortal'*? The answer to the question is already available in the work of Searle (1969) and later with Vanderveken (1985): the type of the link between a locution and its propositional content is dependent on the type of *illocutionary force* which the performer of the speech act assumes. In the AIF+ ontology, the relation between a locution and its propositional content is hence captured by *illocutionary schemes*. Specific applications of these schemes are then, following the now familiar pattern, represented as YA-nodes, which describe passage between L-nodes ("elements" of argument$_2$) and I-nodes ("elements" of argument$_1$). For example, *Plato says,* *'Socrates is mortal'* is linked to *Socrates is mortal* by a YA-node which is an instance of the "asserting" illucutionary scheme.

A link between and L-node and an I-node is warranted by the constitutive rules for the speech act that is performed. In natural contexts, the most important types of constitutive rules are the preparatory and sincerity rules, for which unfulfillment results in defectiveness of a speech act (Searle and Vanderveken 1985). AIF naturally supports different conceptions of speech acts and of illocutionary force in that it allows for multiple sets of illocutionary schemes (just as it allows for multiple sets of argumentation schemes). As a result, it can represent van Eemeren and Grootendorst's (1984) modifications to Searle's and later, Searle and Venderveken's rules and conditions on speech acts. For example, an assertion may be successful but still defective, if its performer declared what in fact he disbelieves: a locutor may not satisfy constitutive rules and still have a chance to perform a successful speech act, since a receiver may not notice their unfulfillment. Thus, the successful adherence to constitutive rules can be viewed as *presumptions* on the applications of illucutionary schemes and all of the existing AIF machinery handles the representation on attacks on the successful application of illocutionary force.

### 3.3. Calculated properties in the AIF

The language of the AIF+ ontology allows us to "record" arguments of both type 1 and 2 and the links between them. However, arguments based on, for instance, counting, weighing, comparing or evaluating other arguments all involve processes (counting, weighing, comparing, evaluating) that cannot be captured in the AIF itself (and nor should they be, for otherwise the AIF would swell to some general purpose programming language). These various processes might collectively be thought of as ways of *calculating properties* about the arguments that the AIF+ ontology represents. It is not that such arguments cannot be represented at all. But rather, if arguments are based on these calculated properties – arguments such as "the prosecution has not provided sufficient evidence for a conviction, so the accused is released" – then they can only be represented in the same way as normal propositions, i.e., as I-nodes. The language of the AIF+ ontology has no way of capturing the link between such a statement and, say, the existence or non-existence of a set of other nodes. For argument$_1$ structures this is a relatively small problem, but excludes, as the previous example demonstrates, some relatively common forms of legal argument. But for dialogue, the matter is more serious. Protocol rules are very often defined on the basis of calculated properties of dialogue histories: the existence or non-existence of particular claims, the current status of claims and commitments.

## 4. Critical discussion in the AIF

Having introduced the pragma-dialectical model of a critical discussion in paragraph 2 and the AIF in paragraph 3, we turn our attention to the correspondence between the two in paragraph 4. We will begin by relating the core concepts of the pragma-dialectical model to the building blocks of the AIF ontology. After which we will tentatively re-introduce the model of a critical discussion in terms of a dialogue protocol by means of a flow-chart that visualises the moves discussants can make within a discussion game and we will highlight some of the most noteworthy and interesting locution pairs found within the protocol.

### 4.1. Pragma-dialectical notions in AIF terms

Evaluating argumentative discourse in accordance with the standard pragma-dialectical model of a critical discussion requires the constructing of an analytic overview. (van Eemeren and Grootendorst 2004, pp. 118–122) This overview covers all analytically relevant, argumentative elements of the discourse. Sections 4.1.1 to 4.1.6 correlate these core elements of pragma-dialectical analysis to the core ontology of the AIF.

#### 4.1.1. Standpoints

In pragma-dialectical theory, a standpoint is a combination of a proposition and an attitude towards that proposition. Clearly, the propositional content of a standpoint corresponds very closely to an I-node in the AIF, but I-nodes (necessarily) omit agent-relativised attitudes towards their content, so an I-node capturing some proposition $p$ cannot directly correspond to a standpoint such as $+/p$. Houtlosser (1994) elucidates the pragma-dialectical foundation that suggests a central role for speech acts, and intimates that offering a standpoint is a distinct speech act, albeit one that may be performed simultaneously with others. We might call the illocutionary force that accompanies such a speech act (rather cumbersomely), 'standpointing'. Armed with this type of illocutionary force, we have a further point of correspondence: a propositional report of a discourse event such as *Bob says p is the case* is captured by an L-node; its propositional content, $p$, is captured by an I-node, and the connection between them is captured by a YA scheme instantiating an illocutionary scheme for standpointing. Bearing in mind that the AIF can directly represent the underlying 'Sentence-level' assertion that also connects the L and I nodes, the picture is as in Figure 3, below.

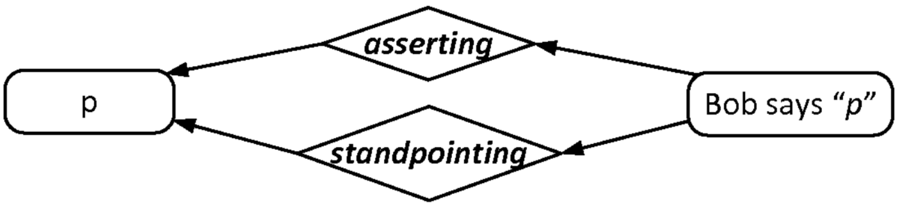



**Figure 3. Standpointing as illocutionary force**

Whilst Figure 3 represents a reasonable AIF interpretation of the speech act constitution of standpoints, it fails to provide us with the locus of a standpoint – although we have a representation of standpointing, we do not yet have one for a standpoint. Two observations lead to a solution. The first observation is that van Eemeren and Grootendort (1984) provide a propositional interpretation of a standpoint, *viz.* (in our example): *Bob's point of view in respect of the expressed opinion* p *is that this expressed opinion* p *is (not) the case* (1984: 114). The second is that this proposition can be deduced from an AIF graph in which there is a sentence level assertion and a higher textual level speech act of standpointing between a single L node and a single I node. In other words, the standpoint can indeed be represented as an I node (it is, after all, a proposition like any other), but one which is a calculated property.

This characterisation of the speech-act nature of standpoints does have some limitations. For van Eemeren and Grootendorst, the relationship between the speech act of standpointing and the speech act of asserting is one of supervention, that is, the content of the standpointing act is precisely the asserting act. The AIF, however, enforces strict type limitations, and is founded upon the early speech act model in which all speech acts (if they have any substantive content at all) have propositional content. As speech acts themselves are not propositions, for the AIF, the passage of illocutionary force captured by the illocutionary scheme cannot itself be the subject of illocutionary force. In this way the current ontology of the AIF prohibits direct connection from one illocutionary scheme to another. Exploring this restriction further in response to the pragma-dialectical approach is an interesting avenue for further investigation.

On the other hand the analysis also has some strengths. The AIF interpretation can cope with Houtlosser's reconstruction of arbitrary speech acts (not just assertives) between the propositional report of the discourse event and the propositional content (i.e. the content of the standpoint), and can similarly handle multiple such speech acts if, for example, both a directive and a (reconstructed) assertive are identifiable at the sentence level. The AIF interpretation also preserves a clear distinction between a standpoint and other speech acts, which is important for subsequent dialogical mechanics (see Section 4.3). And finally, it is possible to expand the analysis presented in Figure 3 explicitly to capture Houtlosser's (1994) more refined account of the complex speech act of standpointing in which it is the acceptability of the sentence level assertive which is the target of the speech act. Illocutionary schemes capture presumptions and constitutive requirements on speech acts in the same way that argumentation schemes capture presumptions and constitutive requirements on inferences. In addition to Searle-like conditions and consitutive rules, the illocutionary scheme for asserting might also typically capture the implicit presumption of acceptability generated by the Interaction Principle. These implicit components act as potential growth points for argument and can be made explicit when appropriate. We could thus revise the picture as in Figure 4, which makes explicit the proposition corresponding to the presumption of acceptability, and then renders that presumption the target of the illocutionary force of standpointing.

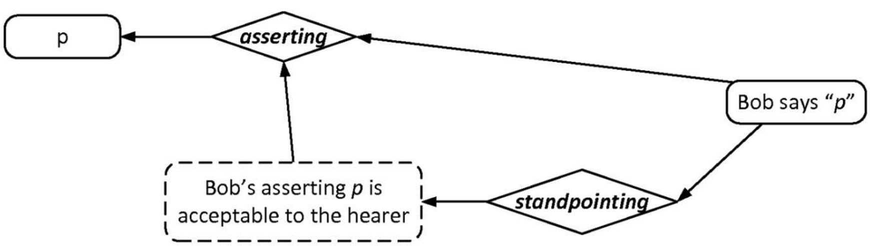



**Figure 4. Standpointing with acceptability of sentence level assertion**

Figure 4 is a significantly more complex interpretation, so for the sake of clarity in what follows, we retain the analysis in Figure 3, because nothing is lost in our investigation if we do so.

#### 4.1.2. Discussion roles

The distribution of the discussion roles is externalised in the opening stage. The discussion parties mutually commit to the distribution for the remainder of the discussion. From then on, every L-node is marked with a specific *agent* property corresponding to a unique name for an interlocutor, and the mapping between these unique names and their roles in this particular dialogue is handled by the commitments established during the opening stage. Thus, for example, we might imagine a move *m* in a dialogue which requires the protagonist to have earlier said *x*. We may have a representation of the utterance of *x* for which the agent property is *Bob*, and furthermore, we may have the parties having committed that for this dialogue Bob is protagonist. The precondition on the move *m* would thus express that there exists some agent about whom there exists a commitment of taking on the role of protagonist, and that this agent must be the value of the agent property of an L-node earlier in this same dialogue.

#### 4.1.3. Starting points

The starting points of an argument are the conceded propositions mutually agreed upon as a part of the common ground, as checked in the intersubjective identification procedure. For the AIF, starting points are represented as I-nodes (starting points in pragma-dialectical theory do not include derivations or applications of inferences, or instances of conflict relations, and so do not include complexes of I-nodes and S-nodes). In pragma-dialectical theory, starting points may also include rules of inference, which correspond to components of the Forms ontology (referred to as F-nodes in (Rahwan *et al.* 2007)). Direct reference to F-nodes from within instances of AIF graphs is not currently possible: it is not possible to argue about or agree to or talk about general rules of inference, as it is in some other systems – particularly those with a legal heritage where the evolution of legal rules is of central importance. This is a known limitation of the AIF which is under investigation elsewhere. Here we limit ourselves to handling propositional starting points. Clearly the propositions that are the subject of the starting points are I-nodes. However, the fact that they are starting points needs to be handled explicitly too. As with much of pragma-dialectical theory, the establishment of starting points has a dialogical basis. As such, the fact that a given proposition is a starting point in a given dialogue is a commitment – that is, an I-node corresponding to a property calculated on the basis of a (set of) L-node(s). So for example, the two L nodes, *Bob said that he thought they both agreed on* p, and *Wilma said that she agreed*, might be used to calculate the property that p *is a starting point*, which itself would be represented as an I-node.

**4.1.4. Argumentation**

The concept of 'an argumentation' in pragma-dialectical theory corresponds fairly closely to O'Keefe's (1977) characterisation of argument$_1$. As a result, an argumentation is simply any connected subgraph of an AIF graph which does not include applications of transitional (TA) or illocutionary (YA) schemes. To include TAs or YAs would be to include dialogue as such, so they must be excluded. Notice however that the definition does allow L-nodes. This is because L-nodes can be used to play a role in arguments$_1$. For example, one might use the premise, *Bob said bananas are yellow* as a basis for an inference to the conclusion that *Bob can speak*, or *Bob knows English*, or *Bob has seen a banana*, and so on. In fact, one rather common use of L-nodes in this way is in arguments from authority (and related forms) – so we must not prohibit L-nodes from appearing in argumentation.

**4.1.5. Argumentation structures**

The pragma-dialectical model recognizes several distinct structures of argumentation, each of which corresponds directly to particular arrangements or constraints on AIF graphs:

- Single argumentation corresponds to a subgraph of AIF involving exactly three nodes: an I-node corresponding to some proposition *p*, an I-node corresponding to some proposition q, and an RA-node connecting *q* to *p*, with the further constraint that there are no other incoming RA-nodes to *p* (in fact this last constraint is rather more difficult to determine since it is relativised to the current dialogue – clearly there might be many other arguments for *p*, but their existence is of no import if they are not adduced in the dialogue at hand).
- Multiple argumentation corresponds to a subgraph of AIF involving at least five nodes: an I-node corresponding to some proposition *p*, two further I-nodes corresponding to propositions *q* and *r*, and two RA-nodes, one connnecting *q* to *p*, the other connecting *r* to *p*. There may be any number of other RA- and I-nodes in the subgraph in addition: the structure described is sufficient for the subgraph to count (at least) as multiple argumentation structure.
- Coordinative argumentation corresponds to a subgraph of AIF involving at least four nodes: an I-node corresponding to some proposition *p*, two further I-nodes corresponding to propositions *q* and *r*, and an RA-node which connects *q* and *r* to *p*. There may be any number of other RA- and I-nodes in the subgraph in addition: the structure described is sufficient for the subgraph to count (at least) as coordinative argumentation structure.

- Subordinative argumentation corresponds to a subgraph of AIF involving at least five nodes: three I-nodes corresponding to propositions $p$, $q$ and $r$, and two RA-nodes, the first connecting $q$ to $p$, and the second connecting $r$ to $q$. There may be any number of other RA- and I-nodes in the subgraph in addition: the structure described is sufficient for the subgraph to count (at least) as subordinative argumentation structure.

**4.1.6. Argument schemes and critical questions**

Argument schemes in pragma-dialectical theory have a direct counterpart in the AIF's representation of rules of inference. The schemes themselves a characterised abstractly (that is to say, uninstantiated) in the Forms ontology, and are then instantiated by RA schemes in specific examples. For the AIF it is important to distinguish the form of, say, Argument from Authority (which defines the form that its premises and conclusion take; defines its presumptions and exceptions; and defines its critical questions), from a given instance of Argument from Authority (which has specific premises, conclusions and possibly some of the implicit presumptions and exceptions made explicit, and possibly some of the critical questions asked).

The pragma-dialectical scheme set, summarised in (van Eemeren *et al* 2002) as comprised of symptomatic, causal and analogical schemes can be represented in the AIF Forms ontology in the usual way, with instances fulfilling the constraints and properties of those forms as with other schemesets already characterised, including those based on Walton *et al.*'s work (2008). Instances of schemes are captured by RA-nodes, and the critical questions correspond, as they do with schemes from other sources, to a variety of structural patterns including implicit premises (I-nodes) for presumptions, implicit conflicts (I-node plus CA-node) for exceptions, and implicit undercutters (I-node plus CA-node plus I-node complex): Rahwan *et al.* (2007) offer some examples of these patterns.

Critical questions form a key part of the machinery of argumentation schemes, and the dual argument$_1$/argument$_2$ nature of schemes and critical questions has been remarked upon previously (Reed and Walton 2007). On the one-hand, schemes and the presumptions and exceptions that the critical questions embody have a distinctly argument$_1$ character, in that they structure the connections between argument$_1$ components. On the other hand, critical questions are inherently argument$_2$ as they need to be asked in order to 'fire'. According to the pragma-dialectical theory, the asking of critical questions is controlled by an intersubjective procedure. Though the results of that procedure correspond to RA nodes and their connected I-nodes, the procedure itself is a part of the dialogical process of critical discussion – in just the same way that Reed and Walton (2007) advocate including a 'Pose' move into a simple dialogue game in order to accommodate the posing of critical questions. It is to the characterisation of these dialogical issues that we turn next.

### 4.2. Towards a critical discussion dialogue game protocol

Drawn from the fifteen rules for a critical discussion and the speech acts that may (or should) be performed by interlocutors in the four stages of a critical discussion, we can characterise the routes along which a dialectical exchange can develop. These possible routes are visualised as a directed graph (or flow-chart) in Figure 5. The discussants start out at the top with one party advancing a standpoint in the confrontation stage. Following the ideal procedure of a critical discussion the discussants can take various routes by performing certain speech acts at specific points during the discussion to move through the opening and argumentation stages and end up in the concluding stage at the bottom of the graph. Momentarily we will treat the intersubjective procedures as 'black boxes', leaving it to the discretion of the discussants to determine the process therein and outcome thereof. These intersubjective procedures are shown as oval nodes in the graph. As is indicated in section 2.4 pragma-dialectical theory does provide insight into these procedures and adding them will be one of the next tasks in the venture of correlating the pragma-dialectical framework to the AIF.

Another proviso we need to make is that in our current tentative take we do not distinguish between the discussion roles and the parties that initially advance a standpoint or doubt it. Remember that either the proponent of the standpoint or the challenger can assume the role of protagonist (or antagonist) in the discussion stage, but ordinarily it will be the proponent of the standpoint who will actually argue for it. Another assumption we make is that the standpoint is positive (i.e. +/p) and is only faced with doubt, not with a contradictory stance. If the challenger would actually take the opposite standpoint instead of merely doubting it, two separate discussions will have to be completed in order to test both the positive standpoint (+/p) and the negative one (−/p). This will solicit a problem of order for the discussants who will have to agree which of the two discussion they will engage in first – and should not be taken as a problem of choice where settling the one dispute would automatically settle the other.[15] At present this fork in

[15] Remember that a standpoint can only be constructively defended. Cf. (van Eemeren and Grootendorst 2004, p. 141) for the problem of *order* (not *choice*) in a mixed or multiple difference of opinion.

the confrontation stage of the discussion has not been incorporated into the flow-chart visualisation of the protocol yet. Catering the protocol for a negative standpoint would be done by allowing for a substitution of the current positive standpoint (+/p) with a negative standpoint (–/p) and requiring the force of the argumentation not to be justificatory for the standpoint but rather refutatory. For the sake of simplicity we will nonetheless stick to characterising a single non-mixed difference of opinion in which a positive standpoint is at issue. Similarly we assume the discussants have no problem understanding each other's utterances and therefore have no need for performing or requesting usage declaratives – which the rules for a critical discussion do allow at any moment (see rule 15 in (van Eemeren and Grootendorst 2004, p. 157).)

Each node in Figure 5 represents a locution performed as indicated by parties 1 or 2 or by both and with its particular discursive function. The edges between nodes represent routes that discussants may take. The first two moves in the discussion will be party 1 advancing a standpoint which allows party 2 to respond to it by casting doubt. Of course in actual discourse interlocutors have the opportunity to perform many more locutionary acts than those shown here. The protocol expressed through the chart only and exactly covers the locutions and locution-pairs which are argumentatively relevant for the dialectical procedure of the critical discussion.[16] Any digression from this procedure will be irrelevant to reasonably resolving the difference of opinion and is not part of the critical discussion procedure. That is to say the protocol presented is normative. For example the discussion party 1 has the possibility to not advance any argumentation and retract his prior standpoint (eg. for the sake of being done with it.) This could be regarded as a move heading directly to the mutual decision to terminate the discussion at the bottom of the flow-chart. But as the discussants did not 'play by the rules' of a critical discussion this path has not been incorporated into the protocol. Such a move would mean there never was a critical discussion to begin with: the standpoint's merits were never put to the test.

A possible difficulty in the procedure represented in the protocol is the move from the antagonists's challenge to either the intersubjective inference or the explicitation procedure. As it stands the first route has to be taken iff the argumentation was both fully externalised and dependent on logical validity in its potential transfer of the acceptability of the premises employed

[16] With the current exclusion of the usage declarations allowed by rule 15 in an attempt to maintain a more-or-less comprehensive chart.

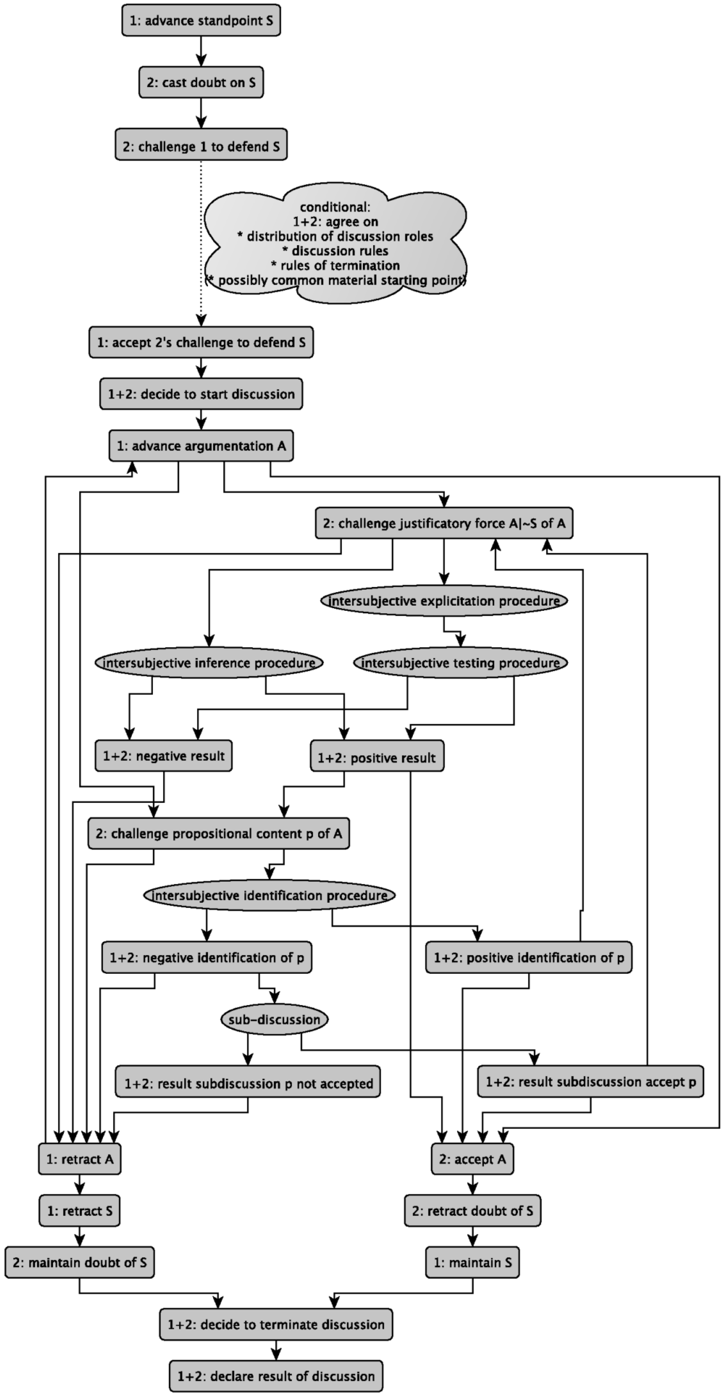



**Figure 5. The (simplified) dialogue protocol of a critical discussion as flow-chart**

to the standpoint. This approach appears to be in line with (van Eemeren and Grootendorst 2004, pp. 148–50). Nonetheless when we regard the reconstructive and interpretative steps available in the analysis of argumentative discourse it appears to be possible to evaluate argumentation that is not presented as fully externalised on the basis of its logical validity. In the absence of pragmatic factors that would suggest otherwise, the analyst can use the transformation method of addition to add any unexpressed premise(s). Such maximally reasonable (charitable) reconstruction is justified because the interlocutors are taken to be bound by the communicative principle of co-operation therefore performing speech acts aimed at the goal of resolving the difference of opinion. (van Eemeren and Grootendorst 2004, pp. 115–118)

The protocol could be amended accordingly (i.e. by allowing the path [challenge force of argumentation] – [intersubjective explicitation procedure] – [intersubjective inference procedure].) This would warrant the question whether in the light of the recent developments in non-monotonic and defeasible logics the strict separation between the intersubjective inference and testing procedures is still viable or even necessary. Although from the perspective of computational complexity inefficient the current protocol allows for the same functionality as a more elaborated procedure which takes the analysis through the logical minimum – the bare-bones needed for coherent inference – and pragmatic optimum – dressing the bare-bones to account for the contextual discursive embedding – into account. An interlocutor could retract argumentation which has failed on the inference procedure side in order to subsequently re-advance it either in fully externalised form or as not based on logical validity this time around. Similarly an analyst can use this method to end up with a fitting maximally reasonable evaluation.

### 4.3 The protocol and locution-pairs in the AIF

By regarding the pragma-dialectical discussion procedure as a dialogue protocol Figure 5 shows that locutions come in pairs where the first might be followed up by one specific or possibly of choice of several successor locutions. Some of the pairs are of more interest than others and by means of example we will characterise six of them in terms of Transition Schemes in the AIF+ ontology. As a result of the universality of the language of the AIF, some of the intricacies of the pragma-dialectical speech acts and critical discussion model need to be treated as calculated properties or left out altogether. Subsequent studies could investigate these omissions further to attempt a more precise correspondence.

A discussion starts with one of the parties advancing a standpoint. The other party may then accept the standpoint, in which case there will be no critical discussion. The more interesting thing to do, from an argumentative perspective, is to doubt the standpoint. Consider the two locutions *advance standpoint* and *cast doubt* in Table 4. In our characterisation of the locutions, we present them in a "semi-formal" way. In Table 4, $p_i$ stands for the party (or player if we see the discussion as a dialogue game) that advances the locution.

| *Locution name* | **Advance standpoint** | **Cast doubt on standpoint** |
|---|---|---|
| *Format* | $p_i$: standpoint $S$ | $p_i$: doubt $S$ |
| *Precondition description* | The propositional content p of *S*...<br>– ...is not in the common starting points<br>– ...has not been the content of another standpoint *S'* in the same discussion | none |
| *Postcond. description* | $p_i$ is committed to (defend) $S$ | none |

**Table 4**

So in order to advance a standpoint, the propositional content p of the standpoint cannot be in the common starting points because the standpoint should in principle not be regarded as fully acceptable (or accepted for that matter) by the other (cf. van Eemeren and Grootendorst 2004, p. 191, commandment 2).[17] Furthermore, advancing a standpoint commits the party to defend this standpoint. (Houtlosser 1994) Notice that there are no pre- and postconditions on the individual *cast doubt* locution. Rather, any conditions on this locution are part of the transition scheme; the characterisation in Table 4 leads us to the first pair of locutions in a discussion that can be modelled as such a scheme in AIF+, viz. Table 5.

| *Scheme name* | **Advance Standpoint → Cast doubt on standpoint** |
|---|---|
| *Format* | $p_i$: standpoint $S \longrightarrow p_j$: doubt $S$ |
| *Presump. description* | $p_i \neq p_j$ |

**Table 5**

Notice that this way of characterising the transition prevents the straw man fallacy by requiring that the standpoint doubted *S* is the same as the one advanced *S* (cf. van Eemeren and Grootendorst 1992, p. 124–31). The transitional scheme adds to this the presumption that the doubt is cast by

[17] Precondition (b) will be discussed below.

a different discussion party from that which advanced the standpoint to account for the dialectical approach.

After casting doubt on a standpoint, there is essentially one possible locution, namely to challenge the party who advanced the standpoint to defend it, viz. Table 6.

| *Scheme name* | **Cast doubt on standpoint → Challenge to defend** |
|---|---|
| *Format* | $p_i$: doubt $S \longrightarrow p_i$: challenge_defend $S$ |

**Table 6**

After a challenge, the other party may accept the challenge or the parties may attempt to set the limits of their discussion by establishing the procedural rules for the discussion, common starting points, discussion roles and termination criteria. On this subject, the literature is somewhat ambiguous: whilst (van Eemeren and Grootendorst 1984, p. 99) seems to indicate that *first* the challenge is accepted and *then* the limits are set, pragma-dialectical rules 3 and 5 above (taken from (van Eemeren and Grootendorst 2004)) state that one is obliged to accept a challenge *unless* there is no agreement on the limits of the discussion. This would indicate that one only has to accept after the limits have been agreed upon. Our (pragmatic) solution is placing the discussion about common starting points and procedures in a meta-dialogue, as is indicated by the cloud-like part in Figure 5. However, in order to stay true to rule 3 we will presume that this discussion has taken place and there is an agreement before someone accepts the challenge to defend a standpoint. In other words, the agreement is a presumption in the transition from challenge to acceptance (Table 7). A further presumption is that the challenger is a different person from the one who accepts.

| *Scheme name* | **Challenge to defend → Accept challenge to defend** |
|---|---|
| *Format* | $p_i$: challenge_defend $S \longrightarrow p_j$: accept_challenge_defend $S$ |
| *Presump. description* | – $p_i \neq p_j$<br>– agreement on discussion roles and rules, starting points and termination criteria |

**Table 7**

Note that the obligation created by the acceptance of the challenge does not have to be explicitly rendered as, for example, a postcondition on the accept challenge to defend locution, as the protocol ensures that the player who accepts the challenge also advances an argumentation in favour of his standpoint: from the *accept challenge to defend* locution it is only possible to go to the *decide to start* locution (Figure 5) and after this locution, there is only one possibility, namely for the party defending the standpoint to *advance an argumentation* in favour of it (Table 8).

| | |
|---|---|
| *Scheme name* | **Advance standpoint & Decide to start → Advance argumentation** |
| *Format* | $p_i$: standpoint $S$ **and** $p_i$, $p_j$: decide_start → $p_i$: argue A |
| *Presump. description* | – $p_i \neq p_j$<br>– if $S$ = (+/p), then A $\mid\!\sim$ p<br>– if $S$ = (–/p), then A $\mid\!\sim$ ¬ p<br>here, $\mid\!\sim$ means so much as "p follows from A" under the agreed rules. |

**Table 8**

Notice that here, there are two locution types that are related to the *advance argumentation* locution. The *decide to start advance argumentation* transition simply denotes the sequence in which the locutions may be uttered: one cannot advance an argumentation before deciding to start a discussion. The relation between *advance standpoint* and *advance argumentation*, however, is a functional (in this case argumentative) one: the argumentation A has to be a reason for or against the propositional content of the standpoint, depending on whether the standpoint is positive or negative.

Note that an *advance argumentation* move can also follow a *retract argumentation* locution. This means that there is another transition to *advance argumentation* viz. Table 9.

| | |
|---|---|
| *Scheme name* | **Advance standpoint & Retract argumentation → Advance argumentation** |
| *Format* | $p_i$: standpoint $S$ **and** $p_i$ retract A → $p_i$: argue B |
| *Presump. description* | – if $S$ = (+/p), then B $\mid\!\sim$ p<br>– if $S$ = (–/p), then B $\mid\!\sim$ ¬p |

**Table 9**

The presumptions of this scheme are slightly different because the utterer of the *retract* and *advance* locutions is the same person.

So now we have a few different conditions on the *advance argumentation* locution: it can only follow a *decide to start* or a *retract argumentation* move, and it has to be in favour of one's standpoint. In a pragma-dialectical discussion, there is another condition on *advance argumentation*, namely that the argumentation has not been advanced yet in this discussion (van

Eemeren and Grootendorst 2004, p. 153). This cannot be modelled as a presumption in the transition scheme in Table 9 (e.g. B $\neq$ A), because the fact that the argumentation has not been advanced before does not just refer back to the just-retracted argumentation advanced immediately before the new one, but rather to all the argumentations advanced in the discussion so far. Something like "all the argumentations advanced in the discussion so far" is a typical example of a calculated property, which is represented in the AIF as a simple I-node, in this case a precondition on the *advance argumentation* locution.

| *Locution name* | **Advance argumentation** |
|---|---|
| *Format* | $p_i$: argue A |
| *Precond. description* | A has not been advanced in this discussion before. |
| *Postcond. description* | $p_i$ is committed to (defend) A |

**Table 10**

After an argumentation has been advanced, the antogonist can either accept the argumentation or challenge the argumentation in various ways (Figure 5). Hence, there are a number of transition schemes from *advance argumentation* to the various challenges, Table 11 and 12.

| *Scheme name* | **Advance argumentation $\longrightarrow$ Challenge propositional content** |
|---|---|
| *Format* | $p_i$: argue A $\longrightarrow$ $p_j$: challenge p |
| *Presumption description* | – $p_i \neq p_j$<br>– p is in A |

**Table 11**

| *Scheme name* | **Advance standpoint & advance argumentation $\longrightarrow$ Challenge justificatory force** |
|---|---|
| *Format* | $p_i$: standpoint $S$ **and** $p_i$: argue A $\longrightarrow$ $p_j$: challenge A $\mid\!\sim S$ |
| *Presump. description* | $p_i \neq p_j$ |

**Table 12**

As discussed in section 4.2, the intersubjective procedures will be left implicit in the current protocol. This means that there are no proper transitional schemes going out from the *challenge argumentation* locutions. The next explicit locution is either a positive or a negative result regarding the justificatory force or a positive or negative identification of the propositional content (Figure 5). Now, in the case of a *negative identification of p*, a sub-discussion is started. This means that there is a transition of the type *negative identification of p → advance standpoint p*. Here, the presumption is that the limits set in the opening stage of the main discussion persevere in the sub-discussion. Also, each proposition p can be debated only once. Similar as for argumentations, this is determined in the preconditions because it refers to a calculated property; recall condition (b) for the *advance standpoint* locution, which says that each proposition p can only be advanced as a standpoint once in a discussion.

In the case of a positive result or identification, we can accept the argumentation or challenge some other part of the argument (e.g. the justificatory force if a proposition p was just positively identified). However, it is important that each proposition can only be questioned once and for any argument can only be questioned once. Again, this can only be modelled as preconditions on the *challenge* locutions: each *challenge* locution has as a precondition that the exact same challenge was not made before during the discussion.

A discussion can only stop if the protagonist retracts his argumentation and subsequently his standpoint or if the antagonist after accepting the argumentation retracts his doubt. Important is that the *retract argumentation A* and *retract standpoint S* have as postconditions that the party that retracts them is no longer committed to (defend) A or S, respectively.

The list of emerging locution pairs and their full specification in terms of presumptions and the locutions' individual pre- and post-conditions is by no means complete. We do believe such a full specification could be construed at a later moment. Let us first summarise what we have done in this paper before returning to the possibilities of continuing the current project.

## 5. Conclusion

Perhaps somewhat surprisingly, pragma-dialectical theory has, by and large, not been taken up in artificial intelligence, due largely to its heavy emphasis on the linguistic and pragmatic structures in natural texts which are extremely challenging for computational accounts to handle. With the advent of the Argument Interchange Format and its focus on representation of real arguments and therefore on pragmatic and illocutionary facets of argumentative discourse, connections between computational models and the pragma-dialectical approach are becoming possible in a more detailed and thorough way than has previously been possible. This paper has taken some initial steps to show how those connections can be made.

In particular, our aim has been to show both the extent and the limitations of computational modelling of the foundational concepts within the pragma-dialectical theory, including standpoints, discussion roles, starting points, argumentation structures and argument schemes. With this basis in place, we have then been able to demonstrate how the complex and sophisticated dialogue game of critical discussion can start to be modelled computationally in terms of the locution types and transitions between locutions, and how that dialogue game can be connected to the underlying argument$_1$ structures that are created, navigated and manipulated by those locutions and transitions. This connection is coherent in both computational AIF terms and also in pragma-dialectical terms. What has been achieved here is just a starting point: much remains to be done both in extending the AIF in the face of representational challenges posed by the pragma-dialectical approach (in terms of illocutionary characterisation of argumentative speech acts, for example), and in continuing to build the connection between AIF and the pragma-dialectical model (in terms of the transitions in the game of critical discussion, for example). An exciting avenue for further investigation is then opened up in being able to explore computationally more recent advances in pragma-dialectical theory such as strategic manoeuvring (van Eemeren 2010). But this paper already demonstrates the potential and the value – for both artificial intelligence and philosophy – of building a computational understanding of the pragma-dialectical approach.

## Acknowledgements

Reed gratefully acknowledges the support of the EPSRC in the UK under grant EP/G060347.

## References

Chesñevar, C., Mcginnis, J., Modgil, S., Rahwan, I., Simari, G., South, M., Vreeswijk, G., and Willmott, S. (2007). Towards an Argument Interchange Format. *Knowledge Engineering Review* 21(4), 293–316.

van Eemeren, F. H., and Grootendorst, R. (1984). *Speech Acts in Argumentative Discussions: A Theoretical Model for the Analysis of Discussions Directed towards Solving Conflicts of Opinion.* Dordrecht: Foris.

van Eemeren, F. H., and Grootendorst, R. (1992). *Argumentation, communication and fallacies: A pragma-dialectical perspective*. Routledge.

van Eemeren, F. H., Grootendorst, R., Jackson, S., and Jacobs, S. (1993). *Reconstructing argumentative discourse*. Tuscaloosa: The University of Alabama Press.

van Eemeren, F. H., Grootendorst, R., and Snoeck Henkemans, A. F. (2002). *Argumentation: Analysis, evaluation, presentation*. Mahwah, NJ: Lawrence Erlbaum.

van Eemeren, F. H., and Grootendorst, R. (2004). *A Systematic Theory of Argumentation: The pragma-dialectical approach*. Cambridge: Cambridge University Press.

van Eemeren, F. H., Houtlosser, P., and Snoeck Henkemans, A. F. (Eds.). (2007). *Argumentative Indicators in Discourse: A Pragma-Dialectical Study*. Dordrecht: Springer.

van Eemeren, F. H., Garssen, B. J., and Meuffels, H. L. M. (2009). *Fallacies and Judgments of Reasonableness: Empirical Research Concerning the Pragma-Dialectical Discussion Rules*. Dordrecht: Springer.

van Eemeren, F. H. (2010). *Strategic Maneuvering in Argumentative Discourse*. Amsterdam: John Benjamins.

Houtlosser, P. (1994). "The speech act 'advancing a standpoint"'. In van Eemeren, F. H., and Grootendorst, R. (Eds.): *Studies in Pragma-Dialectics* (pp. 165–71). Amsterdam: Sic Sat.

Mackenzie, J. D. (1979). Question-begging in Non-cumulative Systems. *Journal of Philosophical Logic* 8, 117–133.

McBurney, P., and Parsons, S. (2009). Dialogue Games for Agent Argumentation. In Rahwan, I., and Simari, G. R. (Eds.) (2009). *Argumentation in Artificial Intelligence* (pp. 261–80). Dordrecht: Springer.

O'Keefe, D. J. (1977). Two Concepts of Argument. *The Journal of the American Forensic Association* 8(3), 121–128.

Prakken, H. (2005). Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation* 15, 1009–1040.

Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation* 1, 93–124.

Rahwan, I., Zablith, F., and Reed, C. (2007). Laying the foundations for a World Wide Argument Web. *Artificial Intelligence* 171 (10–15), 897–921.

Rahwan, I., Banihashemi, B., Reed, C., Walton, D., and Abdallah, S. (2010). Representing and Classifying Arguments on the Semantic Web. *To appear in Knowledge Engineering Review.*

Reed, C., and Rowe, G. (2004). Araucaria: Software for Argument Analysis, Diagramming and Representation. *International Journal of AI Tools* 13(4), 961–980.

Reed, C., and Walton, D. (2007). "Argumentation Schemes in Dialogue". In Hansen, H. V., Tindale, C. W., Johnson, R. H., and Blair, J. A. (Eds.), *Dissensus and the Search for Common Ground* (Proceedings of OSSA 2007), Windsor, ON.

Reed, C., Wells, S., Devereux, J., and Rowe, G. (2008). AIF+: Dialogue in the Argument Interchange Format. *Proceedings of COMMA 2008*, 311–323, Amsterdam: IOS Press.

Reed, C., Wells, S., Budzyńska, K., and Devereux, J. (2010). Building arguments with argumentation: the role of illocutionary force in computational models of argument. *Proceedings of COMMA 2010* (pp. 415–426), Amsterdam: IOS Press.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

Searle, J. R., and Vanderveken, D. (1985). *Foundations of Illocutionary Logic.* Cambridge: Cambridge University Press.

Walton, D., Reed, C., and Macagno, F. (2008). *Argumentation Schemes.* Cambridge: Cambridge University Press.

Jacky Visser
Department of Speech Communication,
Argumentation Theory and Philosophy
University of Amsterdam, The Netherlands
& International Learned Institute for Argumentation Studies
Spuistraat 134, 1012 VB Amsterdam
J.C.Visser@uva.nl

Floris Bex
Argumentation Research Group
School of Computing
University of Dundee
Dundee DD1 4HN Scotland, UK
florisbex@computing.dundee.ac.uk

Chris Reed
Argumentation Research Group
School of Computing
University of Dundee
Dundee DD1 4HN Scotland, UK
chris@computing.dundee.ac.uk

Bart Garssen
Department of Speech Communication,
Argumentation Theory and Philosophy
University of Amsterdam, The Netherlands
& International Learned Institute for Argumentation Studies
Spuistraat 134, 1012 VB Amsterdam
B.J.Garssen@uva.nl

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 23 (36) 2011

**Marcin Lewiński**
Universidade Nova de Lisboa

# DIALECTICAL TRADE-OFFS IN THE DESIGN OF PROTOCOLS FOR COMPUTER-MEDIATED DELIBERATION

**Abstract**: Ideal models of dialectical argumentation, such the pragma-dialectical critical discussion or Walton and Krabbe's persuasion dialogues, comprise of a set of rules that define reasonable argumentation under idealised conditions. Assuming such conditions, dialectical rules are meant to secure an orderly procedure for testing opinions. However, in actual circumstances violations of argumentative rules – identified as fallacies – can and do occur. Pragma-dialectics treats fallacies as "derailments of strategic manoeuvring", that is, contraventions of dialectical rules for a critical discussion committed by actual arguers for rhetorical gains. Hence, the predicament of actual argumentation is a possible (but not necessary) trade-off between dialectical constraints and rhetorical opportunities. In this paper I preliminarily conceptualise a different predicament that actual arguers may face. The sets of dialectical rules proposed in ideal models of argumentation are consistent and thus unproblematic, as long as they presuppose idealised conditions. However, when put to work in actual procedures for argumentation, the rules may clash with one another. For instance, the freedom to unlimitedly criticise the opponent may hinder the progress towards rational resolution of a difference of opinion. As a result, arguers may face a predicament in which the only way to observe one of the rules of reasonable argumentation is to violate another one. I call such possible clashes *dialectical trade-offs*, because they are clashes between dialectical rules that arise in actual circumstances of argumentation. Dialectical trade-offs are practical concerns that do not undermine the general composition and usefulness of the ideal models. Yet, they point to a practical difficulty in designing consistent and applicable protocols for reasonable argumentation. I will illustrate this difficulty by contrasting two kinds of protocols for computer-mediated deliberation: Internet forums for informal deliberation and formal models of deliberative dialogues developed within the field of Artificial Intelligence.

**Keywords**: argumentation, Artificial Intelligence, dialectics, Internet discussion forums, online deliberation, pragma-dialectics

## 1. The overlapping fields of argumentation theory and computer science

Before discussing any topic pertaining to the relation between argumentation theory and computer science, a brief clarification of the senses in which both these fields can be understood is needed. Ever since Aristotle's work on analytics, topics and rhetoric, argumentation theory has traditionally been divided into three sub-fields: logic, dialectics, and rheto-

ISBN 978–83–7431–305–6 ISSN 0860-150X

ric. Wenzel, in his widely recognised account (1979, 1990), proposes to treat logic, dialectics, and rhetoric as three distinct, though interrelated, "perspectives on argument". Logic, in its various forms, focuses on argumentation understood as a *product*, that is, as a certain constellation of sets of claims (premises and conclusions) that lends itself to a precise, formal analysis and evaluation in terms of validity of inferences. Dialectics primarily approaches argumentation as a *procedure* which allows for a critical testing of the disputed opinions. Finally, the rhetorical perspective investigates argumentation as a situated *process* of persuasive communication, often taking place by means of embellished language.

So delineated, the field of argumentation theory seems to be overlapping with computer science first and foremost in its logical and dialectical facet. Computers, in their basic function as computing machines, seem to be predestined to be of great use in supporting methods of abstract, monological reasoning. Proof theories, abstract models of argumentation, and other sub-fields of formal logic and semantics can thus prominently benefit from the opportunities given by computer systems (see Rahwan & Simari, 2009, Ch. 2–12). Some even go as far as characterising the whole area of argumentation in Artificial Intelligence (AI) as "logic continued by other means" (van Benthem, 2009, p. vii). Still, as soon as one focuses on argumentation as a communicative act – rather than purely monological act of internal reasoning – dialectical investigations come to the fore. As Wenzel characterises it, "the dialectical perspective embraces all methodological, procedural approaches to organizing argumentative discussions. The focus of this perspective [...] is on rules, standards, attitudes and behaviors that promote critical decision-making" (1990, p. 16). The purpose of dialectical inquiry is thus to analyse, evaluate, and develop procedures for argumentative discussions: "[w]ithin the dialectical perspective, the chief resources of interest are designs or plans for conducting critical discussions" (Wenzel, 1990, p. 21). It is exactly the issue of designability of argumentative interactions that strongly connects dialectical studies of argumentation to computer science. As has been noticed, new Information and Communication Technologies (ICT) have offered "massively expanded opportunities for deliberate design" and thus precipitated the "explosion of interest in the structure, organization, and conditioning of discourse" (Aakhus & Jackson, 2005, p. 411). The notion of "design" thus links the two fields: the theory and practice of designing computer systems and applications can be utilised for the purpose of creating and testing procedures for argumentative discussions. In this sense, computer science provides the tools for genuinely exercising dialectics in action.

In the use of information technologies for communicative purposes, including argumentation, three kinds of research areas can be distinguished: computer-computer interaction (see Rahwan & Simari, 2009, Ch. 13–17), human-computer interaction (see Sears & Jacko, 2008), and computer-mediated communication between humans (see Herring, 2010). Researchers working in the first of these areas properly belong to the fast-growing field of AI, in that they model possible communication among automated software agents. Such models are based on highly formalised, programmable protocols that make use of many logical insights. However, much work in AI involves design of critical argumentative exchanges (McBurney, Hitchcock, & Parsons, 2007; Prakken, 2000, 2009) and, in this sense, it also functions as an extension of dialectics, especially formal dialectics (Barth & Krabbe, 1982; Walton & Krabbe, 1995, Ch. 4). When it comes to human-computer interaction: while there are hardly any interesting developments in such areas as argumentation between humans and, say, ATM or Xerox machines (nor should there be!), some borderline cases have been of great interest to argumentation scholars. Different kinds of computer supported collaborative work (CSCW), such as group decision support systems (GDSS), in which human agents interact through a computer system have been analysed in terms of their capabilities for supporting dialectically sound procedures of argumentation (Aakhus, 2002a; Karacapilidis & Papadias, 2001; de Moor & Aakhus, 2006; Rehg, McBurney, & Parsons, 2005). Systems for CSCW, however, can also be classified as instances of human-to-human, but computer-mediated communication. The bulk of computer-mediated communication has become part of ordinary people's everyday experience: e-mails, text messages, instant messengers, blogs, online chats and discussion forums, and so on, are commonly used beyond the limited scope of professional activities. While interiorised in a daily experience of millions of users, ordinary formats of computer-mediated communication have noticeable features that can be understood and analysed as constraints on and opportunities for communication and argumentation in particular (Aakhus, 2002b; Jackson, 1998; Lewiński, 2010a, 2010b). Similarly to any other type, or genre, of human communication, various modes of computer-mediated communication have been an object of rhetorical studies (e.g. Benson, 1996; Gurak & Antonijevic, 2009). Still, as protocols for argumentative discussions with explicit design features, they lend themselves particularly well to a dialectical analysis.[1]

[1] Rehg et al. (2005, pp. 209–210) distinguish between different "system roles" in computer-aided procedures for argumentation; crucially, they differentiate between sys-

The goal of this paper is to identify one possibly prominent element of such analysis, which I call a *dialectical trade-off*. Dialectical trade-offs are clashes between different dialectical rules stipulated in the ideal models of argumentation, that arise in actual circumstances. In particular, actual designs of argumentative discussions, whether highly formalised or largely informal, may impeccably embody some of the ideal dialectical rules, yet do so at the expense of other rules. Employing the theoretical apparatus of the pragma-dialectical theory of argumentation, I will argue that this is clear in many computerised protocols for argumentative discussions. Dialectical trade-offs leave the designers and users of such protocols in a predicament that differs from the predicament of actual arguers identified in the pragma-dialectical concept of strategic manoeuvring. Whereas a main source of concern for pragma-dialecticians is a possible clash between "good" dialectics and "bad" rhetoric, I discuss the disconcerting clash between good dialectics and good dialectics. I do so in three basic steps. In section 2, I present the basics of the pragma-dialectical ideal model of a critical discussion and the notion of derailments of strategic manoeuvring. In section 3, I describe the notion of dialectical trade-offs. Finally, in section 4, I identify possible dialectical trade-offs in the design of some computerised designs for argumentative discussions.

## 2. Derailments of strategic manoeuvring: the choice between the good and the bad

The normative, dialectical core of the pragma-dialectical theory of argumentation is embodied in the model of an ideal, dialectical procedure called a *critical discussion* (van Eemeren & Grootendorst, 1984, 2004). A critical discussion is aimed at resolving the difference of opinion between arguers regarding a standpoint by means of a critical testing of the standpoint on the merits. The model specifies the stages, speech acts, and the rules governing the performance of speech acts that are conducive to critically resolving differences of opinion. Pragma-dialecticians formulate the norms of a critical discussion either as 15 rules defined in speech act terms (van Eemeren & Grootendorst, 2004, Ch. 6), or in a simplified form as a "code of conduct" for reasonable discussions, consisting of "the ten command-

tems as supports for human argumentation and systems as non-human participants to argumentation in their own right. Karacapilidis and Papadias clearly state that in the former role "by supporting and not replacing human judgment, the system comes in second and the users first" (2001, p. 260).

ments" instrumental in detecting fallacies of argumentation (van Eemeren & Grootendorst, 1992; 2004, Ch. 8). The rules, in both variants, stipulate basic rights and obligations of the parties to a dialectical discussion: the protagonist and the antagonist. According to the rules, any kind of a standpoint can be advanced and challenged, and, once challenged, it has to be defended on the basis of the procedural and material starting points agreed between the arguers. Further, the rules stipulate what counts as a sound argument (validated by a formal logical inference rule or an informal argument scheme) and a relevant criticism, as well as, ultimately, when a successful defence and attack of a standpoint occurs. In this way, the rules of a critical discussion define reasonable argumentation under idealised conditions (so called "higher-order conditions"). These conditions include free and open participation, equality among arguers, unlimited time for discussion, open access to resources (such as factual knowledge), as well as the cooperative and critical attitude of arguers, who should be persuaded solely by the force of the better argument, rather than private interests and prejudices (van Eemeren, Grootendorst, Jackson, & Jacobs, 1993, pp. 30–34). Assuming these conditions, the dialectical rules of a critical discussion are meant to secure a reasonable and orderly procedure for testing opinions.

It is clear, though, that in actual circumstances violations of argumentative rules can and do occur. Dialectical approaches identify such violations as fallacies of argumentation. The pragma-dialectical theory proposes to treat them as "derailments of strategic manoeuvring", that is, as contraventions of dialectical rules committed by actual arguers for rhetorical gains (van Eemeren & Houtlosser, 2003). The notion of *strategic manoeuvring* is meant to grasp an important predicament that ordinary language users face in their day-to-day argumentation. On the one hand, every serious argumentation by definition involves certain commitment to reasonableness.[2] It means that those who advance a certain standpoint and *argue* for it in order to convince critics – rather than to win them over by various tricks and stratagems, including manipulation, threat and ridicule – make claim to certain standards of reasonableness. These standards, in the pragma-dialectical theory, are embodied in the model of a critical discussion whose norms exactly prescribe what it means to convince a "reasonable critic" by the force of the

[2] Pragma-dialectics defines argumentation as a "verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint" (van Eemeren & Grootendorst, 2004, p. 1). Functionally speaking, crucial in this definition is the formulation of the goal of argumentation as convincing someone who is necessarily "reasonable" and, at the same time, is a "critic" which, until convinced, remains in disagreement over the standpoint proffered.

better argument. On the other hand, also by definition, argumentation is meaningfully advanced only under conditions of a difference of opinion – and resolving this difference in one's own favour is a goal of every genuine arguer. Taking this element into account means that ordinary arguers are most faithfully approached if seen as involved in an agonistic struggle in which victory or defeat are at stake, rather than as purely rational minds aimed at a disinterested quest for truth. What matters in ordinary argumentative discussions is thus a skilful reconciliation of one's desire to win a discussion by getting one's standpoint accepted by the opponent, with the public expectation to do so in a reasonable way. These two distinct considerations may be neatly reconcilable (after all, there is nothing intrinsically wrong in being successful and persuasive), but they may also diverge – in which case a certain tension in pursuing both goals simultaneously may arise. This tension may be resolved by skilful strategic manoeuvring, in which the dialectical norms and rhetorical opportunities are "delicately balanced" (van Eemeren & Houtlosser, 2002). Yet, it may also lead to "derailments of strategic manoeuvring", in which the opportunistic and at times unreasonable rhetorical aspect takes the upper hand over dialectical requirements (van Eemeren & Houtlosser, 2003).

To conclude, pragma-dialectics envisages a predicament of actual argumentation as a possible (but not necessary) trade-off between dialectical constraints of good, reasonable argumentation and rhetorical opportunities that may lead arguers astray to bad, fallacious practices: we all want to persuade others, and sometimes we do so by becoming "too rhetorical", while losing sight of rational limits.

## 3. Dialectical trade-offs: the choice between the good and the good

In this section, I preliminarily conceptualise a different predicament that actual arguers may face. The sets of dialectical rules proposed in ideal models of argumentation, such as the pragma-dialectical critical discussion, are consistent and thus unproblematic, as long as they as they presuppose the idealised conditions mentioned above. However, when put to work in actual, less than ideal procedures for argumentation, the rules may clash with one another. Consequently, a dialectical trade-off may arise.

A dialectical trade-off is a predicament of arguers in actual communicative activities that amounts to a situation in which they are expected to simultaneously observe two (or more) dialectical norms that can hardly be observed together in concrete circumstances. Thus a dialectical trade-off occurs between the rules of reasonable argumentation, but because of actual circumstances. In this sense, dialectical trade-offs are different from the predicament grasped in the notion of strategic manoeuvring, according to which ordinary arguers face a conflict of demands between dialectical rules and rhetorical devices (including fallacious ones) that facilitate successful persuasion. From a normative dialectical perspective, reasonable and responsible arguers should always solve the dilemma of strategic manoeuvring by opting for the dialectically correct solution, even at the cost of rhetorical success. By contrast, dialectical trade-offs present arguers with a more challenging dilemma, in which only one course of action can be taken, and the choice has to be made between following one principle of reasonableness against another. Moreover, the study of strategic manoeuvring has focused on the achievements and limitations of actors trying to solve the conflict of demands for themselves in a rhetorical situation. The way I define dialectical trade-offs does not exclusively point to problems to be solved through individual choices of arguers, but also to dilemmas inherent to the procedural designs of various argumentative activity types (see below, section 4).

These somewhat abstract considerations can be illustrated with the help of a few examples. Jacobs (2003) analyses the tension that may arise in argumentative discussions between "freedom of participation" (that is, access to forums of public argumentation) and "freedom of inquiry" (that is, opportunity to be involved in argumentation of high epistemic value). Both these kinds of freedom are dialectically significant and thus can be analysed as "two values of openness in argumentation theory" (*op. cit.*). According to Jacobs, under ideal circumstances they "converge and complement one another": broader participation leads to a more thorough and critical public scrutiny, while epistemic openness allows for a wide array of opinions to be heard and thus stimulates active participation of all parties concerned (Jacobs, 2003, p. 554). However, constraints of actual circumstances can result in a tension between the two values. For example, time limits in broadcast media can lead to uneasy trade-offs in allocating air-time to various speakers: editors of TV debates can either maximise participation of ordinary members of audience or maximise "expert contributions" (Jacobs, 2003, p. 555). Both of these are valuable from an argumentative perspective, but cannot be achieved at the same time.

Rather more abstract, theoretical instances of possible dialectical trade-offs have been analysed by Krabbe. One of them is the problem of retraction of commitments in dialectical discussions (Krabbe, 2001; Walton & Krabbe, 1995), another, the interrelated issue of conclusiveness of argumentative procedures (Krabbe, 2007). Krabbe's dilemma is the following:

In constructing a precise, formalised model of a dialogue, such as a critical discussion (Walton and Krabbe use the term "persuasion dialogue"), one has to decide to what extent the previously incurred commitments of arguers are retractable. On the one hand, "retractions are needed to bring critical discussions to a successful conclusion" (Krabbe, 2001, p. 144), since a conclusion can only be reached after one of the arguers admits that her position was untenable, and thus retracts it. On the other hand, "if someone keeps denying propositions just asserted, or keeps hedging or evading commitment whenever it appears [...] you can never get anywhere with this arguer" (Walton & Krabbe, 1995, p. 10), since there is no base on which the challenged position can be objected to. To tackle this dilemma, Walton and Krabbe (1995, Ch. 4) develop two formal systems of dialogue rules: Permissive Persuasion Dialogue and Rigorous Persuasion Dialogue. The former is permissive in that it allows retractions, with special conditions attached; the latter is rigorous, as retractions are limited to an absolute minimum (one can, and indeed has to, retract the defeated position). Shortly, "permissive dialogues can be soft on retraction, whereas more rigorous ones need to stick to stricter rules" (Krabbe, 2001, p. 146).[3] So constructed, permissive dialogue accentuates the maieutic function of critical dialectical exchanges. Its chief goal is to comprehensively explore argumentation for and against a disputed position, and by doing so reveal unexpressed premises and other deeply seated, yet implicit, commitments (or, as Walton and Krabbe call them, "dark-side commitments"). Factors of time efficiency and tangibility of outcomes are less of an issue in realising the maieutic task. By contrast, rigorous dialogue functions primarily as a quick and efficient method of deciding whose (sub-)position holds on the basis of the current state of explicit dialectical commitments. Again, both these are dialectically relevant outcomes that often cannot be reached at the same time.[4]

A similar dilemma in the construction of critical dialectical procedures has been analysed by Krabbe (2007) in terms of "predicaments of the concluding stage" of a pragma-dialectical critical discussion. He expounds this predicament as follows:

---

[3] By virtue of Rule 12, the pragma-dialectical critical discussion seems to belong to the permissive type of dialectical dialogues: "The protagonist retains throughout the entire discussion the right to retract any complex speech act of argumentation that he has performed, and thereby to remove the obligation to defend it" (van Eemeren & Grootendorst, 2004, p. 153). All the same, arguers cannot deny or retract premises that belong to the set of accepted starting points.

[4] Walton and Krabbe's solution to this dilemma is to build a theoretical model in which a rigorous dialogue is "embedded into" a permissive one (1995, pp. 163–166). This, however, is a theoretical solution that cannot always solve problems of actual argumentative discussions.

> An argument may seem conclusive, but then the antagonist may come up with new doubt and call into question an element that was previously thought to be uncontested. Similarly, a seemingly conclusive attack may be undercut when the protagonist suddenly sees a new possibility for argumentative defense. Thus there is not much conclusiveness about attacks and defenses being or not being conclusive as long as some party can still add some contribution. (Krabbe, 2007, p. 9)[5]

What Krabbe views as a potential impediment to concluding critical discussions, is at the same time defined in the ideal model as a critically reasonable "optimal use of the right to attack and defend" (see rules 10 and 11 of a critical discussion: van Eemeren & Grootendorst, 2004, pp. 151–152). Making use of these two rights, Krabbe suggests, may thus go against an orderly progress towards a reasonable resolution. Similarly to above, Krabbe's point is a theoretical one. Yet, it becomes of practical importance as soon as one wants to implement an actual procedure for argumentative discussion. Should one aim at an efficient resolution of a difference of opinion and thus prevent such inconclusiveness by putting in place rigid procedural rules and time constraints? Or rather promote an optimal expansion of the disagreement space in which every possible argument and criticism can be voiced without much concern for a timely completion of a procedure?

## 4. Loose protocols vs. formal systems for computer-aided argumentation

Dialectical trade-offs, or dilemmas, discussed in the previous section come to light in the design of procedures for actual argumentative discussions taking place in less than ideal conditions. Features of design of argumentative activities, be them televised public debates, parliamentary discussions, or medical consultations, present arguers with unique opportunities for having reasonable, dialectical encounters. This pertains particularly to computer-mediated formats for argumentative discussions which are explicitly designed through technological choices and thus draw out special at-

[5] However, as Krabbe admits (2007, p. 7), theoretically speaking, in a resolution procedure proposed by van Eemeren and Grootendorst (2004, pp. 143–151), the conclusiveness of argumentation hinges on the successful performance of the *intersubjective procedures*: if the protagonist manages to build his case for the standpoint from arguments checked and accepted in these procedures (i.e., the "fixed inputs", as Krabbe calls them), then he may conclude discussion in his favour.

tention to the ways interaction is mediated. Every argumentation design has a certain – stronger or weaker – preference structure built into it. Such structure makes some options for solving the choice-dilemma related to a given dialectical trade-off more likely to be followed by arguers involved in an activity designed in a particular way.

Researchers interested in investigating the impact of the design of computerised protocols on the shape and quality of argumentation have noted that various systems are successful in fulfilling different argumentatively relevant functions of communication: exploring positions and reasons regarding a disputed issue, coming to a collective decision, or creating a community of experts (Aakhus, 2002a; de Moor & Aakhus, 2006). Every different argumentation design thus accentuates some aspects of argumentative interactions, while downplaying others. For instance, the tools of the systems for "issue-networking" "emphasize opening up lines of argumentation as opposed to closing or limiting lines of argumentation" (Aakhus, 2002a, p. 126). In this way, their design involves a potential dialectical trade-off between extensive and conclusive argumentation.

This trade-off is even more pronounced in various types of informal online discussion forums. Such forums, whether in the guise of Usenet newsgroups or ubiquitous Web-forums, are characterised by a minimal design, which allows for asynchronous and typically threaded discussions taking place by means of messages (posts) similar to e-mails. No formal procedures or strictly formulated and enforced rules govern interactions in such fora. This has a significant impact on the shape of argumentative interactions (Lewiński, 2010a). Here, I focus on but one argumentative aspect of online discussions – their open-endedness.

The rules of most online discussion fora do not support, let alone guarantee, any kind of conclusion to discussions. This means that there is no technologically or institutionally prescribed way of settling disputes in this very type of argumentative activity. Rather, discussions simply fade away without any explicit conclusive results, as soon as users lose their interest in them. Moreover, since online discussions are open-ended, participants to these discussions are by no means obliged, or even expected, to come to any sort of explicitly pronounced decision or agreement on the matters discussed. This is in sharp contrast to the computer-mediated Group Decision Support Systems (Aakhus, 2002a; Karacapilidis & Papadias, 2001; de Moor & Aakhus, 2006; Rehg, McBurney, & Parsons, 2005). As described by Aakhus, the goal of one type of design of such systems is to "funnel" the multi-party argumentative discussion "into a flow from broad differences toward an acceptable conclusion" (de Moor & Aakhus, 2006, p. 97).

Special functionalities – such as evaluating arguments by voting or automated weighing (see Karacapilidis & Papadias, 2001) – are available to users of GDSSs in order to facilitate efficient, time-constrained collective decision-making. Neither such functionalities, nor even time constraints, are programmed into the discussions taking place in informal online fora. As a result, there is always room for a new argument or a new critical reaction to an argument, and any form of coming to a final conclusion may be infinitely postponed. (Of course, discussants may at a certain point explicitly terminate their line of argumentation one way or another, but this would be an empirical incidence rather than an institutional requirement.) The design of such informal fora thus instantiates the theoretical problem raised by Krabbe in his discussion of conditions for concluding critical discussions.

This lack of a tangible, procedurally defined outcome has been viewed as an unwelcome characteristic of online discussions. Critics have pointed out that political discussions which are not explicitly concluded, such as online discussions discussed here, do not bring about any concrete institutional results and are thus futile and pointless. Analysts such as Davis (1999) and Wilhelm (1998) bemoaned online discussions' incapacity to secure "intersubjective agreement" leading to "collective action" (Wilhelm, 1998, p. 316). However, these critics apply a decision-making paradigm or, as they call it, a "problem-solving understanding of conversation [...] geared towards articulation of common ends" (Wilhelm, 1998, p. 329), as a model for evaluation. Yet, this model does not exactly apply to informal political online discussions, since they are not expected to facilitate decision-making but informal opinion-formation (Lewiński, 2010a, Ch. 5). In other words, many political scientists criticise this type of discussions for what they are not meant to be.

From the perspective of the model for a critical discussion one may ask, however, if paradoxically the lack of external pressure on having the discussion ended in a prescribed way and limited time may not enable an extensive critical testing that would approximate the procedures stipulated in a critical discussion? Indeed, as I argued elsewhere (Lewiński, 2010b), informal online forums allow for a thorough public scrutiny of the standpoints advanced. Such scrutiny may be realised through various forms of collective criticism, in which opponents of a given standpoint join forces to critically examine this standpoint.

Taking such considerations into account, one can discern the mechanism of a dialectical trade-off in the design of such informal online discussions. Arguably, they facilitate extensive exploration of the disagreement space, and thus careful critical testing of standpoints and arguments. In particular, when the inventiveness of a single critic of a given standpoint is (temporarily) limited, others may (spontaneously) step in and lend support to her/his criticisms. This, however, may come at a cost of inconclusiveness. Open-ended, pseudonymous, and collective argumentative exchanges are prone to the perils of distributed, or even diluted, responsibility. If anyone can leave a discussion at any time, strict regulation and moderation is lacking, and there is no prescribed way of closing discussions, then no single person may be willing and able to act as an agent able to carry the burden of proof successfully from the confrontation to the point of coming to a reasonable conclusion, and thus to the point at which a difference of opinion is resolved (Lewiński, 2010b). Clearly, then, in terms of dialectical values, the design of informal online discussions tilts towards the unrestricted critical openness of examination among "common people", rather than disciplined, efficient, and conclusive decision-making supported by expert contributions.

Problems of quite the opposite nature are faced by the designers of well-regulated, formalised protocols for computer-aided argumentation and decision-making. McBurney et al. (2007), who develop a formal protocol for an argumentative deliberation dialogue between computerised agents, speak even of an inconsistency (rather than merely a trade-off) between the dialectical requirements that a good, reasonable protocol should meet:[6]

> Essentially, this inconsistency arises because of the need to meet two desirable, but conflicting, objectives in the design of a protocol: freedom for the participants and orderliness of the resulting dialogues. By the very act of defining a protocol for dialogues, we are constraining the freedom of the participants in some way and are imposing some structure on the interactions between them. Because we seek to define a framework within which deliberation dialogues between computational entities can occur, our task, as designers, is to strike an appropriate balance between these conflicting objectives. (McBurney et al., 2007, p. 118)

McBurney and colleagues' method of tackling such "conflicting objectives" is to design a protocol that unequivocally defines the types of allowable and required locutions, as well as sequential rules, yet also opens room for participant's freedom of choice. The latter is difficult to achieve in well-defined,

[6] McBurney, Hitchcock and Parsons (2007, pp. 115–118) detect this inconsistency using the 18 principles of Rational Mutual Inquiry proposed by Hitchcock. While somewhat differently formulated, Hitchcock's principles can easily be rendered in pragma-dialectical terms, and the other way round.

disciplined procedures. One of the elements of freedom that McBurney et al. propose is the possibility to freely retract one's commitments: "we permit participants to make utterances that contradict their own prior utterances, or the utterances of others, and to retract prior utterances" (2007, p. 106). Part of such flexibility, however, can be explained through the nature of the deliberation dialogue, in which participants' (possibly shifting) preferences of one course of action over another may be as crucial as factual knowledge. For related reasons, participants are also free to bring in various issues that they deem relevant in addressing the "governing question" that opens deliberations (*should we take action q?*). Furthermore, following Walton and Krabbe's idea, McBurney et al. extend the freedom of participants by allowing them to embed in their deliberations various other types of dialogue, such as persuasion dialogue, in which proposals, preferences and facts can be tested in a dialectical exchange of pros and cons.

Designers of computerised systems for argumentative discussions thus point to the contradictory problems of over-determination and under-determination. On the one hand, the formal procedures should not be too tightened, for this may stifle free exploration of pros and cons. This is certainly detrimental to any dialectical activity, and may prove disastrous in actual argumentative processes by eliminating some reasonable, yet not so easily supportable options (see Rehg et al., 2005, p. 221). On the other hand, the lack of sufficient regulation can undercut the possibility of reaching tangible results on the basis of orderly exchanges. A solution to this dilemma should be a well-balanced system for argumentative exchanges that stimulates extended critical dialectical exchanges among participants and, at the same time, has clearly defined decision-making capabilities (Karacapilidis & Papadias, 2001, p. 261).[7]

## 5. Conclusion

Dialectical trade-offs described in this paper are practical concerns that do not undermine the general composition and usefulness of the models of idealised dialectical procedures. Rather, the analysis of some dialectical trade-offs points to practical difficulties in designing consistent and

[7] Ultimately, however, the question rests on human factors: "How well participants of diverse backgrounds and capacities make actual use of this formalism in concrete contexts of use remains an open research question to which AI researchers are sensitive" (Rehg at al., 2005, p. 219).

applicable protocols for reasonable argumentation. The notion of dialectical trade-offs seems useful, since it gathers under one conceptual label a set of clearly interrelated challenges that have thus far been treated separately on a case-by-case basis. It may also allow for making a step towards considering certain "priority relations" in incorporating various dialectical rules into actual procedures for argumentation. For instance, in some circumstances, it is more vital to realise the maieutic function of a dialectical encounter by exploring the scope and depth of arguments behind each conflicting position, than to reach a conclusive resolution that would unequivocally promote one position as the most defensible from a dialectical perspective. Conversely, in other circumstances the practical goal of making a reasoned yet timely decision, calls for prioritising the rules that enable an orderly progress of a discussion towards an imminent conclusion. Such distinctions seem to be clear in many commonly experienced argumentative activities. Philosophical and ideological debates do involve thorough exchanges of arguments and criticisms, yet often they cannot be reasonably expected to cease. Legal procedures may take years to conclude, but typically return results that are, arguably, established "beyond reasonable doubt". Practical deliberation leading to a decision regarding the best course of emergency action in case of a fault of a nuclear reactor cannot last longer than minutes, or hours at the most, but still the decision is expected to be taken on reasonable grounds. All such argumentative activities can be, and sometimes explicitly are, supported by means of computerised protocols for discussion. Throughout this continuum of argumentative practices, positions and arguments are advanced and criticised, and thus critical dialectical exchanges are exercised in each case. Yet, different elements of the dialectical system of norms are accentuated or even traded off. Analysts and designers of actual procedures should be sensitive to such differences.

## Acknowledgements

I would like to thank Mark Aakhus for many useful discussions we had on the subject. Some core ideas of this article grew from our collaborative work (Aakhus & Lewiński, 2011). The article is part of a post-doctoral project *Argumentation in the virtual public sphere: Between ideal models and actual practices* sponsored by the Portuguese Foundation for Science and Technology (FCT: SFRH/BPD/74541/2010).

**References**

Aakhus, M. (2002a). Modeling reconstruction in groupware technology. In F. H. van Eemeren (Ed.), *Advances in pragma-dialectics* (pp. 121–136). Amsterdam: SicSat.

Aakhus, M. (2002b). The design of forums for online public deliberation and the consequences for argumentation. *The Kentucky Journal of Communication*, *21*(2), 139–149.

Aakhus, M., & Jackson, S. (2005). Technology, interaction, and design. In K. Fitch & R. Sanders (Eds.), *Handbook of language and social interaction* (pp. 411–436). Mahwah, NJ: Lawrence Erlbaum.

Aakhus M., & Lewiński, M. (2011). Argument analysis in large-scale deliberation. In E. Feteris, B. Garssen, & F. Snoeck Henkemans (Eds.), *Keeping in touch with Pragma-Dialectics* (pp. 165–183). Amsterdam: John Benjamins.

Barth, E. M., & Krabbe, E. C. W. (1982). *From axiom to dialogue: A philosophical study of logics and argumentation.* Berlin: Walter de Gruyter.

Benson, T. W. (1996). Rhetoric, civility, and community: Political debate on computer bulletin boards. *Communication Quarterly*, *44*(3), 359–378.

Benthem, J. van. (2009). Foreword. In I. Rahwan & G. R. Simari (Eds.), *Argumentation in Artificial Intelligence* (pp. vii–viii). Dordrecht: Springer.

Davis, R. (1999). *The Web of politics: The Internet's impact on the American political system.* Oxford: Oxford University Press.

Eemeren, F. H. van, & Grootendorst, R. (1984). *Speech acts in argumentative discussions: A theoretical model for the analysis of discussions directed towards solving conflicts of opinion.* Dordrecht: Foris.

Eemeren, F. H. van, & Grootendorst, R. (1992). *Argumentation, communication, and fallacies: A pragma-dialectical perspective.* Hillsdale, NJ: Lawrence Erlbaum.

Eemeren, F. H. van, & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach.* Cambridge: Cambridge University Press.

Eemeren, F. H. van, Grootendorst, R., Jackson, S., & Jacobs, S. (1993). *Reconstructing argumentative discourse.* Tuscaloosa: University of Alabama Press.

Eemeren, F. H. van, & Houtlosser, P. (2002). Strategic manoeuvring: Maintaining a delicate balance. In F. H. van Eemeren & P. Houtlosser

(Eds.), *Dialectic and rhetoric: The warp and woof of argumentation analysis* (pp. 131–159). Dordrecht: Kluwer.

Eemeren, F. H. van, & Houtlosser, P. (2003). Fallacies as derailments of strategic maneuvering: The *argumentum ad verecundiam*, a case in point. In F. H. van Eemeren, J. A. Blair, C. A. Willard & A. F. Snoeck Henkemans (Eds.), *Proceedings of the fifth conference of the International Society for the Study of Argumentation* (pp. 289–292). Amsterdam: SicSat.

Gurak, L., & Antonijevic, S. (2009) Digital rhetoric and public discourse. In A. A. Lunsford, K. H. Wilson & R. A. Eberly (Eds.), *The handbook of rhetoric* (pp. 497–507). Thousand Oaks: Sage.

Herring, S. C. (2010). Computer-mediated conversation: Introduction and overview. *Language@Internet, 7*, Retrieved March 31, 2011, from http://www.languageatinternet.de/articles/2010/2801/index_html/.

Jackson, S. (1998). Disputation by design. *Argumentation*, *12*(2), 183–198.

Jacobs, S. (2003). Two conceptions of openness in argumentation theory. In F. H. van Eemeren, J. A. Blair, C. A. Willard & A. F. Snoeck Henkemans (Eds.), *Proceedings of the Fifth Conference of the International Society for the Study of Argumentation* (pp. 553–556). Amsterdam: SicSat.

Karacapilidis, N., & Papadias, D. (2001). Computer supported argumentation and collaborative decision making: the HERMES system. *Information Systems*, *26*(4), 259–277.

Krabbe, E. C. W. (2001). The problem of retraction in critical discussion. *Synthese*, *127*(1–2), 141–159.

Krabbe, E. C. W. (2007). Predicaments of the concluding stage. In H. V. Hansen, C. W. Tindale, J. A. Blair, R. H. Johnson & D. M. Godden (Eds.), *Dissensus and the search for common ground* (CD-ROM), Windsor, ON: OSSA.

Lewiński, M., (2010a). *Internet political discussion forums as an argumentative activity type. A pragma-dialectical analysis of online forms of strategic manoeuvring in reacting critically.* Amsterdam: SicSat.

Lewiński, M. (2010b). Collective argumentative criticism in informal online discussion forums. *Argumentation and Advocacy*, *47*(2), 86–105.

McBurney, P., Hitchcock, D., & Parsons, S. (2007). The eightfold way of deliberation dialogue. *International Journal of Intelligent Systems*, *22*(1), 95–132.

Moor, A. de, & Aakhus, M. (2006). Argumentation support: From technologies to tools. *Communications of the ACM*, *49*(3), 93–98.

Prakken, H. (2000). On dialogue systems with speech acts, arguments, and counterarguments. In M. Ojeda-Aciego et al. (Eds.), *Proceedings of JELIA'2000. The 7th European Workshop on Logics in Artificial Intelligence* (pp. 224–238). Berlin: Springer.

Prakken, H. (2009). Models of persuasion dialogue. In I. Rahwan & G. R. Simari (Eds.), *Argumentation in Artificial Intelligence* (pp. 281–300). Dordrecht: Springer.

Rahwan, I., & Simari, G. R. (Eds.) (2009). *Argumentation in Artificial Intelligence.* Dordrecht: Springer.

Rehg, W., McBurney, P., & Parsons, S. (2005). Computer decision-support systems for public argumentation: Assessing deliberative legitimacy. *AI & Society*, *19*(3), 203–229.

Sears, A., & Jacko, J. A. (Eds.) (2008). *The Human-Computer Interaction handbook: Fundamentals, evolving technologies and emerging applications* (2nd ed.). New York: Lawrence Erlbaum.

Walton, D. N., & Krabbe, E. C. W. (1995). *Commitment in dialogue: Basic concepts of interpersonal reasoning.* Albany: State University of New York Press.

Wenzel, J. W. (1979). Jürgen Habermas and the dialectical perspective on argumentation. *Journal of the American Forensic Association*, *16*, 83–94.

Wenzel, J. W. (1990). Three perspectives on argument: Rhetoric, dialectic, logic. In J. Schuetz & R. Trapp (Eds.), *Perspectives on argumentation: Essays in honor of Wayne Brockriede* (pp. 9–26). Prospect Heights, IL: Waveland.

Wilhelm, A. G. (1998). Virtual sounding boards: How deliberative is on-line political discussion? *Information, Communication & Society*, *1*(3), 313–338.

Marcin Lewiński
ArgLab, Institute of Philosophy of Language(IFL), FCSH
Universidade Nova de Lisboa
Avenida de Berna 26–C
1069–061 Lisbon, Portugal
m.lewinski@fcsh.unl.pt

**Karolina Stefanowicz**
University of Białystok

# NEW TOOLS FOR SOCIAL DIALOGUE ON THE INTERNET. OPPORTUNITIES AND THREATS FOR NEW SOCIAL SPHERE

**Abstract**: New media have always changed the face of the public sphere. Social domain has entered the era of interactive tools that have revolutionized the existing functions of the Internet. This article is an analysis of new social and communication situations on the Internet that modify already existing communication processes. It aims to present new forms of public sphere and their possible communication advantages and disadvantages. The main focus of the analysis are social media because they are an important novelty in the new networking tools. Furthermore, it seems that social media can contribute to the increasing of the level of democracy and building civil society in countries where citizens are active and are regular users of such media. It does not exclude, however, many situations in which these new tools distort the public sphere and make manipulation more easy than ever before. First part of my observations is based on twentieth century philosophical conceptions of public sphere. Second, characterizes main tools of the new social dialogue and the consequences of its use. Theses, supported by recent examples, point to opportunities and threats of this new phenomenon.

**Keywords**: social sphere, social media, political marketing, manipulation, social dialogue

Nowadays, more and more elements of civic life move to the Internet. Cyberspace is characterized by sophisticated technology commonly referred to as the Web 2.0., which enables its users a wider participation in the virtual world and increases the importance of the content generated by them. Day by day, the Web becomes a more widespread and more interactive tool which redefines the relation between the user and the medium, revolutionizes interpersonal contacts and increases the exchange of information. We are now dealing with mechanisms and applications which enable us to engage in dialogue, exchange opinion, political views or political advertisements – all this based on mutual relation with others. This fundamentally changes the nature of communication, since, until now, only one-sided messages were possible. The Internet creates a space for collective communication and debate. The debate may take different forms: ephemeral exchange of opinions, heated debates, controlled and artificial discussions or casual comments on

ISBN 978–83–7431–305–6 ISSN 0860-150X 

a given subject. Aside from the possibility to communicate, cyberspace also affects other spheres of human life. The Internet becomes the catalyst for transnational social movements, a forum for expressing opposition or disapproval, a place to incite to civil disobedience or to have an informal discussion on social problems. The new, virtual world has become the extension of public sphere to the point that it is sometimes the only place where real public debate and exchange of opinions takes place.

## 1. The XX's century concept of social sphere

The concept of social sphere is inherently included in contemporary theories and definitions of democracy. A short introduction into the origins of 20$^{th}$ century theories of the phenomenon will help us to understand it better. In modern democratic society public sphere is the place for political discourse and exchange of opinions. After the World War II the issue of social dialogue was originally taken up in Western Europe by Hannah Arendt and Karl Popper and in subsequent years by Jurgen Habermas and Ralf Dahrendorf. The philosophy of the two former was influenced by their shocking experiences connected with totalitarian regimes. The evolution of this concept after the Second World War is deeply rooted in the ideas of equality, respect for individual rights and promotion of tolerance. The fearful effects of entrusting small groups or individuals with all power led these thinkers to natural conclusions that authority should be as much dispersed as possible. Public sphere becomes the main tool of control over the authorities and the guarantor of individual freedom. Similar conclusions are reached by twenty years younger Jurgen Habermas and Ralf Dahrendorf, for whom public sphere became the remedy for the experience of war, even though it was not their personal experience. Unique and thorough analysis of public sphere is common for all four authors whose analysis, on the one hand, starts a new trend in social thought, and on the other hand, takes a great deal from the achievements of the past, especially ancient Greece. The new hybrid of ideas puts people on a pedestal, and at the same time places them in a completely new social-political environment. This combination and numerous ideas on how to put it into practice are truly extraordinary.

These concepts are very inspiring even today. They became increasingly popular with the development of the idea of participatory democracy in the West which is based on social dialogue and originated as an answer to many problems and obstacles connected with the functioning of the parliamentary democracy. They were also present on the eastern side of the Iron Curtain, where the opposition first comprised the sphere in the form of underground movement and then emerged to the surface. Manuel Castells writes that from the standpoint of social theory, *space is the material support for contemporary social practices.*[1]

In my article I aim to prove that many social practices, including social dialogue, changed diametrically as a result of the development of the Internet. It is well-justified to claim that the Internet, at its current stage of development has become a vital element of nowadays social sphere. Hannah Arendt, when defining public sphere, says that everything that appears publicly may be seen or heard by everybody and has the widest possible range of recipients. In this sense, the new media: radio, press, television, all essentially fulfil this condition.[2] The second important element of public sphere is the sense of community. It is a sense which is common to us all; however it differs, depending on our personal position in the community.[3] Public sphere is thus a place of activity which goes beyond personal interests of individuals. In this context, the Internet can be treated as a global public arena. Only this medium makes it possible to create and disseminate information, generate virtual social situations and produce interaction and bonds with others. Another argument for the recognition of the Internet as a segment of public sphere may be the one provided by Habermas which states that public sphere is the area of human social activity where public opinion is formed. It is an area where people can freely discuss important social issues and through this discussion influence the actions of politicians.

It may monitor the authorities and create space where individuals gather to discuss problems which are important to the public.

## 2. Development of Internet

The evolution of social space on the Internet progressed gradually. At first, it assumed only the form of posting comments to press articles and participating in newsgroups, however, with the development of new techno-

[1] Castells Manuel. *Społeczeństwo sieci* (*The Rise of the Network Society*). Warszawa 2008: Wydawnictwo Naukowe PWN. p. 412.

[2] Arendt Hannah. *Kondycja ludzka* (*Human Condition*). Warszawa 2010: Wydawnictwo Alteheia, p. 70.

[3] Ibidem, 23.

logies, such as mobile Web, social networking websites, and other communication tools, the debate became much more widespread. The Internet, at first slow and costly, now fast and relatively inexpensive, is slowly reaching to all people around the world, which is well taken care of by both IT companies and the consumers themselves, as they can appreciate the new technological developments.[4]

There are more and more of those who are willing to inhabit the virtual world, and become more and more active in it. Even though the Internet is yet not as widespread as television, in many regions of the world the number of those who are digitally excluded is diminishing. The study by Eurostat (2010) shows that in 2009 fifty nine percent of households in Poland were connected to the Internet.[5] This trend is still growing. New solutions reach especially young people whose attitude towards the Internet is very open and who treat it as the main source of information. Moreover, the number of Internet users is also growing in other age groups, especially among older people, aged 55–65.[6]

The second very important factor, especially for the theory of argumentation is the flattening of the hierarchy of actors involved in the discussion. Nowadays, the boundary is blurred between the sender and the receiver or between authority and the ordinary user of the Internet. Multiplicity of non-hierarchical relationships makes the discussion more open. Arguments become more equivalent, and less supported by the force of authority or the media. Furthermore, the broadcaster can be anyone connected to the Internet. These new opportunities have not been available in such a wide range up to now and are dramatically changing the face of the Internet. New, more egalitarian tools are a key stimulator of social dialogue in cyberspace.

What is more, the Web has overcome many spatial and geographical barriers. A new identity of *the citizen of the world* who participates in global social movements and seeks for internet-based, transnational communities representing his views is being formed. Global identity is no longer a utopian construct, but becomes more and more real and almost every day shows its new face – the face which is changing so fast that it is hard to predict what it will look like even in the near future.

---

[4] Friedman Thomas L. Świat jest płaski. Krótka historia XXI wieku (The World Is Flat: A Brief History of the Twenty-First Century). Poznań 2006: Dom Wydawniczy Rebis. p. 65.

[5] http://www.stat.gov.pl/cps/rde/xbcr/gus/PUBL_nts_spolecz_inform_w_polsce_2006-2010.pdf.

[6] Facebook Data Team. 2011.

## 3. Tools of social dialogue

An important element of the virtual space are social networking sites with their mechanisms and applications. Particularly popular are sites with developed, interactive, and varied applications which combine many different functions. These new mechanisms called social media are focused on social interaction, using highly accessible and scalable communication techniques based on web and mobile technologies. Such websites include, among others, Facebook, Twitter, MySpace, Hi5, and the Polish Blip. Lifestreaming tools allow for information to get into circulation amazingly fast. The world's largest TV channels and newspapers can be found on Twitter. Also, it is becoming one of the main sources of information for journalists.

The most popular Facebook, although it has been operating only since 2004, has 600 million active users, of which 70% live outside the U.S. (Facebook Statistics: 2011). Each of them has an average of 130 "friends". Additionally, the webpage has 160 million entities with which its users can interact. Founded in 2006, Twitter until now gathered up 175 million Internet users, that is 100 million more than last year (January 2011).

Facebook.com links friends, acquaintances, family members and companies with customers. Users can create profiles with photos, lists of hobbies, contact information, and other personal information. Members of website community can exchange information, pictures, documents and links, build networks of interest, use and create tools for communication, fun and games, including automatic notifications when they update their profile. It is possible thanks to basic applications: Photos, Notes, Groups, Events and Posted items (the comments in the form of text, link, or video). Users can also communicate with friends and other users through private or public messages and a chat feature. They can also create and join interest groups and "like pages". The second very popular side, Twitter, on the basis of the so-called microblogging, enables its users to send and read short text messages called *tweets*. Tweets are text-based posts of up to 140 characters displayed on the user's profile page. Notes written by the user profile by logging on the main site, sending an sms or use of third party applications appears in the pages of all other users connected to it.[7]

Enormous numbers of users quickly found various groups which, in many ways, try to use the websites to disseminate information. Facebook was founded in a dormitory by a student Mark Zuckenberg, who, when he

---

[7] Nations Daniel. 2008. *What is a Tweet*: http://webtrends.about.com/od/glossary/g/what-is-a-tweet.htm.

started off with his project, did not expect that it would become so popular or that, apart from "regular users," Facebook would be joined by many institutions being able to take advantage of many useful applications of the website. Originally, it was supposed to serve only social purposes in the academic environment and help with the exchange of information in a faster way. But suddenly, many independent centres gained subjectivity and power to influence others. Besides friends and family, social networking sites gained new type of users: associations, social movements, parties and other political organizations, pressure groups, news programmes, "old media" such as newspapers and television, think-tanks, and even "historical figures"[8] and celebrities. Their impact takes many forms: opinion formation, influencing voters and encouraging to vote, political marketing, encouraging open public debate and contributing to the development of participatory democracy and new political discourse. Mike Westling, University of Washington, writes that there is no other online community that connects members of real-world communities (geographic, ideological, or otherwise) in such an effective way. *Facebook combines the best features of local bulletin-boards, newspapers, and town hall meetings and places them in one location that is available at any time in practically any location.*[9] The great advantage is that, in contrast to the discussions at the Town Hall, Facebook allows all members of the worldwide community to have an input on a topic giving them the flexibility of deciding when and how they contribute to the conversation.

The most common activity on the websites involves creating a network of supporters, e.g.: of a party, movement, reform or idea. Users create thematic groups and introduce topics from many different areas of life. Characteristic of such groups is the fact that every member can invite other users to join the group. In groups, members can participate in discussions, exchange links, and send latest news items. The act of joining a particular group may be a sign of empathy, sympathy, or participation in the group in real life. Moreover, the groups may serve as polling tools, and to some extent replace street surveys. It is more and more common to come across such statements in newspapers: *32 thousand of Facebook users is against... 43 is for...*

---

[8] On Facebook one can become a friend of, e.g. Karl Marx and Józef Piłsudski and a fan of Plato, Che Guevara, Kierkegaard or Ludwig Zamenhof.

[9] Westling Mike. *Expanding the Public Sphere: The Impact of Facebook on Political Communication, 2007*, http://www.thenewvernacular.com/projects/facebook_and_political_communication.pdf 4.01.2011.

### 4. New Opportunities and Threats

Emerging doubts concern the question whether we are dealing with a real social and political discourse. In addition, one cannot yet provide a clear classification of the new phenomena, because they are very dynamic and varied. It is worth stressing that the websites became a place for open and assertive expression of social and political views by private users, which, however, is not often transferred outside the realm of media. Tomasz Mastyk[10] rightly points out that *the Internet and associated technologies are necessary but not sufficient tools for building a strong democracy based on deliberative principles. It is easy, however, taking into account its possibilities, to believe in its power and motive force to influence the form and content of social activity. Still, it must be remembered that it is the only tool that can significantly enhance and promote active citizenship and encourage participation in the process of deliberative-rational political discourse.* According to Bernard Manin and Azi Lev-On,[11] the Internet as a medium is a *mixed blessing.* On the one hand, these groups are an ideal place for like-minded people in which there is a high risk of reducing the debate and ignoring opposing points of view. On the other hand, we can observe groups that invite to an open debate on controversial topics, while at the same time, very few real-world discussions allow opponents to speak.[12]

A specific style of spreading information on the Internet is worth noting. It is significantly different to other media, i.e. television and newspapers. An Internet user who looks for a piece of information chooses himself or herself where to find it and ceases to be a passive consumer.[13] Having this possibility to choose he or she is more willing to trust the information being provided. A wide variety of sources of information to choose from enables him or her to find the websites which will reflect his/her views. This often

---

[10] Masłyk Tomasz. *Demokracje deliberatywna a Internet* (*Deliberative Democracy and the Internet*) in: Magdalena Szpunar (red.). *Media a polityka* (*Media and Politics*). Rzeszów 2007: Wydawnictwo Wyższej Szkoły Informatyki i Zarządzania, p. 265.

[11] Manin Bernard, Lev-On Azi. *Happy accidents: Deliberation and online exposure to opposing views.* in: Todd Davies, Seeta Gangadharan (red.). *Online Deliberation: Design, Research and Practice.* Chicago 2009: Center for the Study of Language and information, p. 113.

[12] Westling Mike. *Expanding the Public Sphere: The Impact of Facebook on Political Communication. 2007.* http://www.thenewvernacular.com/projects/facebook_and_political_communication.pdf 4.01.2011.

[13] Wit Hubert. *Cyfrowe nierówności w dobie kultury 2.0. Problemy wybrane* (*Digital Inequalities in the Age of 2.0. Culture. Selected Issues*) in: M. Jeziński (red.). *Nowe media a media tradycyjne* (*New Media and Traditional Media*). Toruń 2009: Wydawnictwo Adam Marszałek, p. 21.

leads to the situation in which the Internet user restricts himself or herself to those sources of information which match his or her previous viewpoint. This naturally reduces the chance for a change of views. In effect, it leads to the polarization of his or her earlier assumptions and continuous surfing away from the views different from one's own. Hence, increased focus is put on the processes of consolidation of earlier views and grouping with those who support them, rather than a change in views.

In the 1980s, American psychologists described the so-called "hostile media phenomenon". It consists in the fact that groups firmly in support of given ideas perceive a neutral, balanced messages on television as hostile, because the medium did not provide such description of reality which they know is true.[14] On the Internet this phenomenon is not so visible, but this is at the cost of another distortion. New communication tools on the one hand trigger discussion, but on the other hand lead to strong polarization of views. We are free to choose our information providers in such a way that only the content which agrees with our views and expectations gets through to us. This may have very negative effects, previously not encountered on such a scale. A today's Internet user may feel that he or she participates in a public confrontation of views, whereas the percentage of opposing views with which he is confronted is very small. In addition, there exists significant social consent to various forms of manipulation, covert advertising, creating sham grassroots movements and initiatives. In fact, if a person is conscious of such operations he or she may feel confused, and the more he or she is unaware of them the easier it is to manipulate them.

## 5. Political marketing

At a conference organized in 2010 by the Facebook corporation, its Vice President – Blake Chandlee – encouraging participants to advertise on their website – stressed that the marketing research shows that 78% of people trust their friends' recommendations, and only a small percentage of advertising as such.[15] It seems that this applies not only to consumer choices,

[14] Aronson Elliot, Wilson Timothy D., Akert Robin M. *Psychologia społeczna. Serce i umysł* (*Social Psychology. The Heart and The Mind*). Poznań 1997: Wydanctwo Zysk i S-ka, p. 138.

[15] Blake Chandlee is the vice president and commercial director of Facebook, responsible in the company for the actions in Europe, the Middle East and Africa who appeared at a conference called Facebook Now, organized by Facebook and a marketing group ARBOinteractive on 12 January 2010 in Warsaw. The Conference was devoted to substantive issues related to advertising on Facebook.

but also to many others, including the socio-political, and one has to remember that high trust usually means a big political impact. Transferring this to socio-political sphere we can easily notice that news on the Internet spreads like gossip, quickly reaching bigger and bigger audience (snowball effect applies here). The mechanisms of information transmission on the web functions similarly as in private life. Each time we reach for a source it is most often because it was recommended to us by our acquaintances.

Advertisers, who in the search for new ways of reaching their customers take advantage of this mechanism and call it *buzz marketing.*[16] It is a powerful tool to pass on information through informal contacts which gives the illusion that we are dealing with a private opinion. The illusive bond which is formed with the recipient makes manipulation easier.

Politicians create their profiles and websites on the Internet, which enables them to bring their supporters together and to remind them for whom they should vote, to reinforce their views and to build social capital. What is more, public figures such as members of parliament, local politicians, publicists, create the so-called press corners where one can find current information, see latest posts, photos, reports from important events or engage in a virtual conversation with them.[17] Undoubtedly, a big advantage of social media is independence in providing information. The content is often not processed by journalists, but formulated directly by the authors and in this direct form reaches Internet users. *Most of these online forms of communication are used to reduce the distance between a voter and a politician, and increase competitiveness in political communication.*[18]

Those who follow posts on particular websites write comments in which they express their political attitudes. Unfortunately, apart from positive effects such innovations might bring, like better access to information, they also change the image of a politician which starts to remind one of a film star created by image experts. Often, thanks to the comments of Internet users a politician has an opportunity to learn what and how to speak in order to gain widest possible support. This way, however, they fall in a trap

[16] Hughes Mark. *Marketing szeptany. Z ust do ust. Jak robić szum medialny wokół siebie, firmy, produktu* (*Buzzmarketing. Get People to Talk About Your Stuff*). Warszawa 2008: Wydawnictwo MT Biznes.

[17] Leszczuk-Fiedziukiewicz Anna. Kampania negatywna i kampania internetowa jako przykład promocji polityków w kampanii wyborczej do Parlamentu Europejskiego w 2009 roku (Negative Campaign and Internet Campaign As an Example of Promotion of Politicians in Campaign to the European Parliament in 2009.) in: E. Kużelewska, A. R. Bartnicki (red.). Zachód w globalnej i regionalnej polityce międzynarodowej (The West in Global and Regional International Politics). Toruń 2009: Wydawnictwo Adam Marszałek, p. 276.

[18] Ibidem, 256.

of “temporary moods” and, if their main motivation is to please their potential voters, they change into “political aggregates of dispersed comments of Internet users”. Thus, virtual comments may influence political decisions and attitudes in the real world. It is so much dangerous as the Internet by no means reflects the opinion of the whole of society but only some of its segments or even random, individual users who often rebel against the whole socio-political sphere.

People gathered around an association, a politician, or a social movement generate social capital and possess certain participatory potential and therefore there emerges more and more strategies to realize it and use it. In order to maximize the number of recipients, already persuaded “fans” are asked to invite their friends and share information on private profiles. In view of the fact that an average user has a lot of “friends” the new range of influence can be huge, much bigger than with other channels of communication. In real life we share our political views only with good friends and family, whereas such websites become the place of political exhibitionism. In addition, on-line activity often moves to “off-line” life. In Poland, one of the profiles which encourage active participation is: *Powszechna mobilizacja na wybory prezydenckie* (*General activation for presidential election*) and the so called “event” *Ratujmy program TVP Kultura. Zbieramy głosy pod petycją* (*Save the TVP Kultura channel. We are collecting signatures for a petition*). *General activation for presidential election* is a website which works as a newsletter, sending current information to users with the aim to motivate them to take action. In the case of the event there is only a piece of information about an initiative and about ways of supporting it. If somebody wants to sympathize with a chosen movement, they can join it in the virtual space and through this indirectly manifest their view, since the information on the fact of joining a given group may be seen on the profile page of the user and is also sent to all of his or her friends.

The power of the Internet is especially appreciated in the United States. According to the Pew Institute during last year’s campaign, 46% of Americans used the Internet as a source of information about politics, 30 percent watched election spots in the network, and about 10 percent of citizens have used their social networks account for the purposes of the election campaign. Moreover, thousands bloggers were invited to the meetings with Barack Obama or John McCain.[19] Barack Obama’s political promotion is regarded

[19] Trzeciak Sergiusz, *Internetowa demokracja* (*Internet Democracy*), *Tygodnik Onet* 2009, http://www.trzeciak.pl/pl/onet_pl_internetowa_demokracja.html.

to be the best and most effective on the Internet. In addition to political marketing in the Internet, he asked local Internet users for help in organizing the campaign. Building first strong relationship with the users and then making many "activation moves", he managed to stimulate people's commitment and real actions which contributed to his victory. For this purpose he used every opportunity to present himself in a positive light. Americans could be on first-name terms with him on many social networking sites. They could watch his pictures, films showing his political and private life or read his political commentaries. Those who wanted to find out more could visit www.barackobama.com, listen to podcasts[20] with interviews, and subscribe to a mailing list. Currently, Obama's campaign is emulated by many politicians around the world, both in developed democracies and in countries with unstable political mechanisms. Facebook or Twitter host not only politicians from almost every European country, but also candidates in presidential campaign in Sri Lanka or politicians from Rwanda.

## 6. "Social Media Revolution's"

The Internet and social networking sites have become a place of debate on issues which in many ways are ignored by other media and a place where new areas of open political discourse are born. Interactive communities give birth to different social movements. This is, among other factors, due to low costs incurred by senders and recipients, concurrent existence of horizontal and vertical communication, shared responsibility for the content, promotion of equality, the fact that information can spread so fast, and lack of borders. These are undoubtedly the benefits of the Internet for democratic politics[21] which is now strengthened and developed in a completely different direction than before. Thanks to this, one can see more and more manifestations of deliberation which has been taking place for some time now. Already in the early nineties in Santa Monica, California, a heated Internet discussion about the problem of homelessness took place. What is more, the homeless themselves took part in it. Today, such debates are a common occurrence. Online discussions are taking place among members of political and social movements, citizens gathered around particular institutions and

[20] Format of Internet audio recording, allowing quick access to various recordings, including many archival radio broadcasts.

[21] McQuail Denis. *Teoria komunikowania masowego* (*McQuail's Mass Communication Theory*). Warszawa 2007: Wydawnictwo Naukowe PWN, p. 165.

groups of friends. A particular advantage of Facebook is bringing all sources of information to one website. A recent improvement of the social networking website allows the user to use some of his or her favourite websites from one main site, as they are linked to Facebook. The user has all news appearing on his wall. This results in the fact that much more information reaches users than in a situation when they would have to visit each website individually.

Also, completely new fields of discussion emerge. On Facebook or Twitter people take up topics which are absent in other media. On the pages of particular Facebook groups one can read about death penalty in Iran (e.g. *I bet I can find 1,000,000 Against Government Violence in Iran* or *In Defence of Freedom of Speech in Iran*), riots in Palestine or controversial, unofficial opinion on the accident in Smolensk. The discussions are not only among supporters of a view, but also among its opponents, particularly in response to posts of journalists and columnists. It often happens that two people of different views confront accidently, as they happen to have a mutual friend or belong to the same group. New technologies allow for one single post to produce a widespread response among Internet users. A good example here would be a seemingly irrelevant act of a teenage Polish girl who founded a Facebook group called *Mówię stanowcze nie zburzeniu Pałacu Kultury* (*I strongly oppose the demolition of The Palace of Culture and Science in Warsaw*) which, within only two weeks had over 8,500 supporters, and now, after half a year of existence (data from April 2010), over 17,500 thousand. With the fast dissemination of information such groups make a strong activation base and prove that the Internet becomes the main tool of communication.

Even more spectacular and easy-to-notice effects can be observed in such countries as Iran or Moldova. In these two places young people went out into streets encouraged by local activists on western websites. Twitter and Facebook constitute a perfect medium to promote social problems and it was thanks to those websites that in spring 2009 on Teheran's and Chişinău's town squares crowds of young people appeared to protest.[22] In both cases the students used Twitter and Facebook as a forum for fast exchange of messages on computers and mobile phones, which allowed holding demonstrations and informing about their progress. Foreign press described the events, first in

---

[22] *Iran. Wojna w Internecie. Facebook i Twitter za Teheranem* (*Iran. War in the Internet. Facebook and Twitter for Teheran*). 2009.06.18 http://www.redakcja.newsweek.pl/Tekst/Polityka-Polska/530070,Iran-wojna-w-internecie-facebook-i-twitter-za-teheranem.html, 1.12.2010.

Moldova, then in Teheran as *Twitter revolution.*[23] This phenomenon was described in detail by Evgeny Morozow in the *Foreign Policy* magazine. On April 7, 2009 he checked the most popular topic of discussion on Twitter within the past 48 ours. Most people in the global Web observed and wrote posts tagged, rather strangely, as follows: "# *pman* (aside from such tags as *Eminem* or *Easter.*) *Pman* is an acronym of *Piata Marii Adunari Nationale*, the name of the biggest square in Chişinău, the capital city of Moldova, where anti-Russian protests took place [Morozov, 2009]. Even though the political context was different, the events in Iran were similar. After a tragic death of a sixteen-year-old Neda, an Iranian girl shot by the police during an opposition demonstration, thousands of identical posts could be read on Twitter and Facebook: *Rest in peace Neda. The world is crying watching your last breath, but your death shall not be in vain. We will remember you.*[24] This type of publicizing activity by Internet users is not an isolated case, but only one of many more examples of actions carried out on the Internet by young "cyberactivists".

Similar actions are especially typical for non-democratic countries where the Internet is the only place for expressing views and meeting and debating with others, and serves as a tool for activating people. Those countries show the biggest numbers of people who use online sources and Western social networking sites. In Egypt, already in 2008 Imams in mosques preached that using Facebook is a great sin. The biggest, and so far the only action against the Internet in Egypt was a complete blockage of the Web on January 28, 2011. In the middle of the day, during escalating street riots, five biggest Internet providers in the country blocked access to the Internet for all of their users, as probably ordered by the government. Even China, which carefully monitors Internet traffic, did not ever go that far.[25] Events like the ones described above can be observed in many other places. Numerous Twitter accounts disappeared during the "green revolution" in Iran, and currently thousands of Algerians' profiles on Facebook were removed. This shows how regime governments fear the power of the new medium, and at the same time proves that it plays an important role in shaping public opinion.

[23] Morozov Evgeny. *Moldova's Twitter revolution. Foreignpolicy 2009.04.07* http://net-effect.foreignpolicy.com/posts/2009/04/07/moldovas_twitter_revolution 12.12.2010.

[24] *16 letnia Nada. Męczennica irańskiej rewolucji* (*16-years-old Nada. A Martyr of Iranian Revolution*) 2009. [http://wiadomosci.gazeta.pl/Wiadomosci/1,80591,6741791,16_letnia_Neda_meczennica_iranskiej_rewolucji.html, 28.04.2010.

[25] http://wyborcza.pl/1,75477,9025243,Jak_wylaczono_Internet_w_Egipcie_Tego_nie_bylo_nawet.html#ixzz1Cf1i1hvh.

## 7. Conclusions

We are dealing with a variety of phenomena belonging to the positive concept of cyberdemocracy. It is a participatory democracy, in which citizens generate content, participate in virtual communities or even help in Internet-based voluntary organizations. On the other hand, we are entering the world of meticulous control and manipulation of citizens, their consumer and political rights, and control of information which they read or make available on the Internet. Both negative and positive results of the development are to a great extent determined by a cultural, financial and social capital of Internet users and authors. *Communication techniques, by themselves, have no direct impact on the habits and cultural practices.*[26] Depending on social, cultural, and economic factors, political activity on the Internet is realized through many ways of building virtual contacts and forms of online relationships. Users and Internet environment constitute foundations of quasi-society, or even quasi-state – without a real superstructure with which they generate discourse or which they defy. The answer to the question of how the new, interactive possibilities of Web 2.0. will be utilized depends largely on the potential of Internet users – citizens, their ingenuity, but also their ability to act beyond the realm of technoreality. It is not the Internet, as a medium, that creates a new political arena, but people who make use of this tool.

## References

1. Arendt, H. (2010). *Kondycja ludzka* (*Human Condition*). Warszawa: Wydawnictwo Alteheia.
2. Aronson, E., Wilson, T. D., Akert, R. M. (1997). *Psychologia społeczna. Serce i umysł* (*Social Psychology. The Heart and The Mind*). Poznań: Wydanctwo Zysk i S-ka.
3. Castells, M. (2008). *Społeczeństwo sieci* (*The Rise of the Network Society*), Warszawa: Wydawnictwo Naukowe PWN.
4. Friedman, T. L. (2006). *Świat jest płaski. Krótka historia XXI wieku* (*The World Is Flat: A Brief History of the Twenty-First Century*). Poznań: Dom Wydawniczy Rebis.

[26] Goban-Klas Tomasz. *Media i komunikowanie masowe. Teorie i analizy prasy, radia telewizji i Internetu* (*Media and mass communication. Theories and analysis of the press, radio, television and the Internet*). Warszawa 2006: Wydawnictwo Naukowe PWN, p. 143.

5. Goban-Klas, T. (2006). *Media i komunikowanie masowe. Teorie i analizy prasy, radia telewizji i Internetu* (*Media and mass communication. Theories and analysis of the press, radio, television and the Internet*). Warszawa: Wydawnictwo Naukowe PWN.
6. Habermas, J. (2006). Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research. *Communication Theory*, volume 16, issue 4, November 2006.
7. Hughes, M. (2008). *Marketing szeptany. Z ust do ust. Jak robić szum medialny wokół siebie, firmy, produktu* (*Buzzmarketing. Get People to Talk About Your Stuff*). Warszawa: Wydawnictwo MT Biznes.
8. Leszczuk-Fiedziukiewicz, A. (2009). Kampania negatywna i kampania internetowa jako przykład promocji polityków w kampanii wyborczej do Parlamentu Europejskiego w 2009 roku (Negative Campaign and Internet Campaign As an Example of Promotion of Politicians in Campaign to the European Parliament in 2009). In: E. Kużelewska, A. R. Bartnicki (Eds.). *Zachód w globalnej i regionalnej polityce międzynarodowej* (*The West in Global and Regional International Politics*). Toruń: Wydawnictwo Adam Marszałek.
9. Manin, Bernard, Lev-On, Azi (2009). *Happy accidents: Deliberation and online exposure to opposing views.* In: T. Davies, S. Gangadharan (Eds.), *Online Deliberation: Design, Research and Practice.* Chicago: Center for the Study of Language and Information.
10. Masłyk, T. (2007). *Demokracje deliberatywna a Internet* (*Deliberative Democracy and the Internet*). In: M. Szpunar (Ed.), *Media a polityka (Media and Politics).* Rzeszów: Wydawnictwo Wyższej Szkoły Informatyki i Zarządzania.
11. McQuail, D. (2007). *Teoria komunikowania masowego* (*McQuail's Mass Communication Theory*). Warszawa: Wydawnictwo Naukowe PWN.
12. Morozov, E. (2010). *Moldova's Twitter revolution. Foreignpolicy 2009. 04.07* http://neteffect.foreignpolicy.com/posts/2009/04/07/moldovas_twitter_revolution 12.12.2010.
13. Nations, D. (2008). *What is a Tweet*: http://webtrends.about.com/od/glossary/g/what-is-a-tweet.htm.
14. Trzeciak, S. (2009). *Internetowa demokracja* (*Internet Democracy*), *Tygodnik Onet* 2009, http://www.trzeciak.pl/pl/onet_pl_internetowa_demokracja.html.
15. Westling, M. (2007). *Expanding the Public Sphere: The Impact of Facebook on Political Communication*, http://www.thenewvernacular.com/projects/facebook_and_political_communication.pdf 4.01.2011.

16. Wit, H. (2009). *Cyfrowe nierówności w dobie kultury 2.0. Problemy wybrane* (*Digital Inequalities in the Age of 2.0. Culture. Selected Issues*). In: M. Jeziński (Ed.), *Nowe media a media tradycyjne (New Media and Traditional Media*). Toruń: Wydawnictwo Adam Marszałek.
17. *16 letnia Nada. Męczennica irańskiej rewolucji.* 2009. [http://wiadomosci.gazeta.pl/Wiadomosci/1,80591,6741791,16_letnia_Neda_meczennica_iranskiej_rewolucji.html 28.04.2010.
18. *Facebook. Statystyki*, http://www.facebook.com/press/info.php?statistics 11.02.2011.

Karolina Stefanowicz
Department of Political Science, Faculty of Law
Univeristy of Bialystok
ul. Mickiewicza 1, 15–213 Białystok, Poland
www.prawo.uwb.edu.pl
stefanowicz.karolina@gmail.com

## NOTES ON CONTRIBUTORS

**LEILA AMGOUD** received the Computer Science Engineering degree from the National Institute of Computer Science (Algeria) in 1995 and the Ph.D. degree from the Paul Sabatier University (France) in 1999. From 1999 to 2000, L. Amgoud was a Research Assistant at the University of Liverpool. Since 2001, she is a researcher at the French National Center of Scientific Research (CNRS). Presently, she is a Director of Research at IRIT laboratory (France). She is mainly interested in Artificial Intelligence: Non-monotonic reasoning, Argumentation theory, Handling inconsistency in knowledge bases, Reasoning about preferences, Merging knowledge bases, Decision making, Learning, Dialogue and negotiation, Agent architectures and Agent communication languages. L. Amgoud has served on the program committee in several AI conferences and has published more than hundred papers. She is the head of ADRIA (Argumentation, Décison, Raisonnement, Incertitude et Apprentissage) group and an associate editor of the Journal 'Argument and Computation'. Finally, she was involved in several national and European projects on argumentation theory.

**FLORIS BEX** is a postdoctoral researcher working on the Dialectical Argumentation Machines project at the University of Dundee's School of Computing. He has an MSc in Cognitive AI from Utrecht University, and in 2009, he obtained his PhD degree (cum laude) at the Centre for Law and ICT in Groningen (the Netherlands). The subject of his dissertation is how to use hypothetical stories and arguments to make sense of a mass of evidence in a legal (criminal) case; the dissertation was published in early 2011 by Springer as *Arguments, Stories and Evidence – a Formal Hybrid Theory.* Floris' main interests concern computational and philosophical aspects of argumentation in various domains, such as multi-agent systems and legal reasoning. Currently, the focus of his research is on the connection between argumentative dialogue and argument structure, the connections between different (logical and non-logical) argumentation frameworks and the role of narratives in argumentation.

**ZBIGNIEW BONIKOWSKI** is an assistant professor at the Institute of Mathematics and Informatics at the Opole University. He has Ph.D in mathematics in a range of computer science (1996, ICS PAS). His main research include methods of representation of vague, incomplete and uncertain knowledge and reasoning under incomplete and vague information. Particularly his research focus on formal methods and theories for data analysis and inference inspired by concepts of rough set and formal concept. He is a member of the Polish Association for Logic and Philosophy of Science (PTLiFN), the *Polish Philosophical Society* (PTF) and the Group of Logic, Language and Information (GLLI). He was a member of Program Committees of international conferences (IICAI-05, RSKT 2008, 2010, 2011, RST 2011).

**EDWARD BRYNIARSKI** – Ph.D., a lecturer at the Institute of Mathematics and Informatics, Opole University, Poland. Research areas and fields of scientific interests: rough sets theory, logics, philosophical foundations of mathematics and computer science, theoretical foundations of mathematics and computer science, cybernetics and metacybernetics, cybernetic foundations of mathematics, rationality in science, philosophy of science, information and communication technology in elementary and secondary education, computer science education.

**KATARZYNA BUDZYŃSKA** specializes in theories of argumentation, dialogue and persuasion. She was a visiting scholar at universities in the UK (12-month visit at School of Computing, University of Dundee in 2009/10, Ministry of Science and Higher Education Program "Support for International Mobility of Scientists") and in the US (3-month visit in the Center for Cognitive Science, Rutgers University in 2005 and 3-month visit at the University of Michigan in 2004). Budzyńska is an author of over 40 publications. Budzyńska has been invited and funded to give talks and courses at universities in Dundee (UK), Waterloo (Canada) and Girona (Spain). In May 2010, she was invited to participate at a workshop which was a continuation of a project of an international consortium on a standard language of argument description AIF. Budzyńska is a member of program committees for national and international conferences (e.g. *Student Session at the 13th European Agent Systems Summer School*; *Logic and Cognition 2012*) and a member of organizing committee at *14th Workshop on the Semantics and Pragmatics of Dialogue*. She is also a member of the scientific committee at *COGENCY. Journal of Reasoning and Argumentation* published by Diego Portales University (Santiago, Chile). Budzyńska is an assistant professor in the Institute of Philosophy at Cardinal Stefan Wyszyński University in Warsaw.

**FLORENCE DUPIN DE SAINT-CYR** is lecturer in computer science at Toulouse University since April 2005 and member of the ADRIA team (whose key words are Argumentation, Decision, Reasonning, Uncertainty and Learning) of IRIT (Toulouse computer science research institute). She obtained her PhD degree from Toulouse University in December 1996, this same year she was teacher assistant at Reims University. From 1997 to 2003, she was lecturer at Angers University where she was member of the LERIA. Her research field is Artificial Intelligence with a particular focus on the following topics: Reasonning about change (Revision, Updating, Extrapolation) in knowledge base systems, Causality, Argumentation and Dialog, Weighted logics (penalty logic, possibilistic logic), Preference handling. She has been reviewer for, as well as author in, international journals and conferences (such as Artificial Intelligence, Fuzzy Sets and System, AAAI, IJCAI, KR). Finally, she was involved in several national projects on artificial intelligence.

**BART GARSSEN** is Assistant Professor in the Department of Speech Communication, Argumentation Theory and Rhetoric of the University of Amsterdam. He published on argument schemes, fallacies and political argumentation. Recently Garssen published a monograph on laymen's evaluation of fallacies: *Judgments of reasonableness*. Garssen currently studies argumentation in debates in the European Parliament.

**MAGDALENA KACPRZAK** is an assistant professor in Department of Computer Sciences, Polish-Japanese Institute of Information Technology in Warsaw and in Bialystok Technical University (Poland). She holds a Ph.D. in mathematics and computer science from the University of Warsaw. For 11 years her work is concerned in the main with modal logics, and formal methods, with emphasis on the application of these to the specification and verification of cognitive processes in multi-agent systems. Her current research is focused on model checking and parametric verification of properties of persuasive distributed systems. Scientific experience she was derived from cooperation with 13 scientists from 6 countries. She is an author and co-author of many papers in refereed international journals. Full list include over 50 articles. They have 238 citations on Google Scholar. In 2002–2008 she was a member of the VerICS group at the Institute of Computer Science, Polish Academy of Sciences which under the supervision of professor Wojciech Penczek works on developing a model checker tool for automatic verification of timed and multi-agent. Kacprzak is a fellow of Polish Association for Logic and Philosophy of Science. In years 2003–2005 she was a treasurer of this association. Since 2008 she plays an active part in developing ArgDiaP initiative which popularize polish school of argumentation.

**MARCIN KOSZOWY** is assistant professor at the Chair of Logic, Informatics and Philosophy of Science, University of Białystok. He obtained his PhD degree from John Paul II Catholic University of Lublin in 2008. His interests cover argumentation theory, methodology of science and semiotics. His current research is focused on the connection between argument analysis and automated reasoning. Among other things, Koszowy wrote several papers on the concept of information, logical fallacies, and the study of reasoning in the Lvov–Warsaw School. From 2001 to 2008 he co-organized *Stanisław Kamiński Memorial Lectures* (Faculty of Philosophy, John Paul II Catholic University of Lublin). In 2011 he was a member of the Organizing and Programme Committee of the ArgDiaP conference meetings.

**MARCIN LEWIŃSKI** (PhD, University of Amsterdam, 2010) is a post-doctoral researcher at ArgLab, Institute of Philosophy of Language, Universidade Nova de Lisboa, Portugal. His main field of interest lies at the intersection of argumentation theory (in particular pragma-dialectics), political theory, linguistic pragmatics and the analysis of information and communication technologies. He is currently involved in two projects sponsored by the Portuguese Foundation for Science and Technology (FCT): *Argumentation in the virtual public sphere: Between ideal models and actual practices* (FCT-funded individual post-doctoral project) and *Argumentation, Communication and Context* (FCT-funded institutional project in ArgLab). Marcin's main publications include argumentative analysis of discussion groups on the Internet (*Internet political discussion forums as an argumentative activity type*, Amsterdam 2010), fallacy analysis, and study of multi-party argumentation (or argumentative *polylogues*) in the political context.

**PAWEŁ ŁOZIŃSKI** is a Ph.D. student at the Faculty of Electronics and Information Technology, Warsaw University of Technology. His doctoral thesis is devoted to context-dependent, argument-based reasoning. His other research actitities include recent management of a technology-incubation project *IMS Incubator*, participation in a grant for the development of *Multicommodity Market Data Model* and in a grant for *Adaptive problem solving support system based on the analysis of available sources of electronic content.*

**CHRIS REED** is Professor of Computer Science and Philosophy at the University of Dundee in Scotland, where he heads the Argumentation Research Group. Chris has been working at the overlap between argumentation theory and artificial intelligence for over a decade, has won over £1m of funding from RCUK, government and commercial sources and has over 100 peer-reviewed papers in the area including five books. He has also been instrumental in the development of the Argument Interchange Format, an international standard for computational work in the area; he is spear-heading the major engineering effort behind the Argument Web; and he is a founding editor of the Journal of Argument & Computation.

**KAROLINA STEFANOWICZ** is assistant at the Faculty of Law at Univeristy of Bialystok. She has MA in Sociology (2008, Univeristy of Bialystok) and BA in European Studies (2009, Univeristy of Bialystok). Her research interests cover civil society theory, sociology of the Internet and public sphere. At the present moment, she is working on a PhD about NGOs operating in the sphere of the Internet.

**KAZIMIERZ TRZĘSICKI** is the head of the Chair of Logic, Informatics and Philosophy of Science, Univeristy in Białystok. His main subjects of interest are logic and its application in informatics.

**JACKY VISSER** studied in the research master's programme "Rhetoric, Argumentation theory and Philosophy" at the University of Amsterdam. During his studies he was research assistant to Prof. Frans H. van Eemeren and participated in organising the 7th conference of the International Society for the Study of Argumentation (ISSA). In the academic year 2011–2012 he teaches at the University of Amsterdam (department of Speech communication, Argumentation theory and Rhetoric) while preparing a Ph.D. research proposal. Following his background in Cognitive Artificial Intelligence (Utrecht University), his main research interest lies at the intersection of argumentation theory and artificial intelligence. In his master's thesis he investigated a possible formalisation of the pragma-dialectical ideal model of a critical discussion.

**JACEK WALDMAJER** is an assistant professor at the Institute of Philosophy, Opole University. He has MA in mathematics (Inst. of Mathematics, University of Opole) and Ph.D in philosophy (Dept. of Philosophy and Sociology, Maria Curie-Skłodowska University). His research interest covers modal logics (epistemic logic, dynamic logic), multi-agent systems, adequacy of knowledge representation, communication theory, rationality in science and philosophy of science. He is a member of The Association for Symbolic Logic, The European Association for Logic, Language and Information, Polish Association for Logic and Philosophy of Science, *The Polish Philosophical Society,* Group of Logic, Language and Information.

**DOUGLAS WALTON** (PhD University of Toronto, 1972) is a Canadian academic and author, well known for his many widely published books and papers on argumentation, logical fallacies and informal logic. He presently holds the Assumption Chair in Argumentation Studies and is Distinguished Research Fellow of the Centre for Research in Reasoning, Argumentation and Rhetoric (CRRAR) at University of Windsor, Canada. Walton's work has been used to better prepare legal arguments and to help develop artificial intelligence. His books have been translated worldwide, and he attracts students from many countries to study with him. A festschrift honoring his contributions, *Dialectics, Dialogue and Argumentation: An Examination of Douglas Walton's Theories of Reasoning and Argument*, ed. C. Reed and C. W. Tindale, London: College Publications, 2010, shows how his theories are increasingly finding applications in computer science.

**URSZULA WYBRANIEC-SKARDOWSKA** – Ph.D., Habilitation, was for many years a full professor of logic at the Opole University, recently a co-founder and co-chairperson of the Group of Logic, Language and Information established at this academic center. Her interdisciplinary research interests include: logics, philosophy, theory of communication and argumentation, formal linguistics, information sciences and mathematics. In 2000, she was a visiting professor at Tilburg University and University of Amsterdam. In Spring 2009, she was a visiting professor at CUNY (Ph.D. Program in Mathematics) and guest at Stanford University, CA. She is a member of many Polish and international scientific associations, included: The Association for Symbolic Logic (USA), The European Association for Logic, Language and Information (FoLLI), The Polish Division of Logic, Methodology and Philosophy of Sciences of the International Union of History and Philosophy of Sciences, the International Association for semiotic Studies, and the Polish Society of Universalism, the Polish Society of Philosophy (vice-president of Opole Division), the Polish Society of Mathematics.